(Continued)

CONTENTS—*Continued from preceding page*

CONTENTS—*Continued from preceding page*

CONTENTS—*Continued from preceding page*

Time-reversal focusing of elastic surface waves with an asymmetric surface layer

Pelham D. Norville and Waymond R. Scott, Jr.
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332
norville@ece.gatech.edu, waymond.scott@ece.gatech.edu

Abstract: The effectiveness of time-reversal focusing is evaluated in the presence of an asymmetric surface layer that changes the direction of the propagating waves, but does not continually scatter or block the propagating wave front. Interactions between the wave front and the surface layer are dependent on the depth and material properties of the asymmetric surface layer and its orientation in the medium with respect to the incident wave. Time-reversal focusing is shown to perform significantly better than other excitation methods for the purpose of delivering energy to the location of a buried land mine.

1. Introduction

Time-reversal focusing is a powerful technique that allows propagating waves to be focused to a particular location. Time-reversal focusing is most useful when it is difficult or impossible to characterize the clutter and wave propagation speed in an area of examination. While other focusing methods require some knowledge of the propagation medium characteristics such as propagation speed, time reversal does not require this information. Because of its insensitivity to clutter and variations in wave-propagation velocity, elastic-wave time reversal shows great promise for application to the detection of buried objects.

The primary experimental and numerical investigations of time-reversal focusing have been in the ultrasound frequency range and in the far field.[1–3] Time-reversal focusing has been investigated in both fluid and solid media, including liquid-solid interfaces. In fluid media, homogeneous backgrounds have been augmented with scattering objects to create high order scattering of incident waves.[4] In these scenarios, time-reversal focusing has been shown to be effective in inhomogeneous media,[5] even producing super-resolution effects in some cases.[4,6,7] The investigation of time-reversal focusing of Rayleigh waves in elastic solids was first motivated by the detection of surface and subsurface flaws in a solid.[8,9]

These experiments provide a strong foundation in the investigation of acoustic time reversal, but the experimental cases that have been examined are still significantly different from those encountered in the buried object detection problem. For buried land mines, the frequency range of interest is typically centered around 400–1000 Hz, and the detection problem must be carried out in an elastic solid background medium. Inhomogeneity may be encountered by the addition of hard solid objects, such as rocks or other near-surface scatterers, by voids in the medium, or variations in the background material properties.

The work presented in this letter investigates the performance of time-reversal focusing in soils, with a particular emphasis on the detection of buried targets, especially buried land mines, in the presence of an asymmetric surface layer. In field conditions, many types of surface layers may be present between a land mine and excitation sources. Compacted road beds or tire tracks, soil stratification, concrete slabs, and other types of surface layers may significantly alter the propagation direction of any wave front incident on them. In these cases, an excitation wave

Fig. 1. (Color online)The experimental facility. The seismic transducer array is on the right and the ground-contacting sensor array is positioned over the sand tank. The plywood wedge-shaped asymmetric surface layer is buried flush with the surface of the sand.

front may fail to deliver significant energy to the desired location. Time-reversal focusing is an appealing way to deliver energy because it can provide tight and accurate focusing even in the presence of inhomogeneity, such as surface layers.

While time-reversal focusing has been studied extensively in fluids, and even in some solid media, studying the phenomenon in soil entails a significantly different analysis. Only limited-scale studies in any type of granular media have been performed,[10,11] indicating a need for further study. The complexity of wave propagation in soil goes beyond that of many common fluids or elastic media. Soil is a complicated nonlinear particulate medium in which the interparticle interactions along with the specific physical characteristics of the particles define the behavior of waves supported in the medium.[12]

2. Experimental method

2.1 Experimental setup

The experimental results are obtained in a laboratory at the Georgia Institute of Technology (Fig. 1).[11,13] A large concrete wedge-shaped tank is filled with approximately 50 tons of damp compacted sand. Sand is chosen as the background medium because its seismic properties are similar to many types of soil, and because it is straightforward to recondition disturbed sand. This allows for easy burial and removal of scattering objects and targets in the tank.

The seismic waves are generated by an array of six electrodynamic shakers. Each source is made from an 8.7 cm diameter transducer from an Aura bass shaker that has a flat frequency response over the frequency range of interest. A short metal bar foot is attached to each electrodynamic shaker. The shaker and metal foot are placed in contact with the sand and the $12.5 \text{ cm} \times 1.27 \text{ cm} \times 2.54 \text{ cm}$ aluminum bar foot couples seismic energy into the sand.

Once the shakers are used to excite elastic waves in the sand tank, an array of specially designed ground contacting accelerometers[14] is used to record the acceleration of the surface of the ground. These sensors are inexpensive, compact, and couple to the ground lightly enough to be safe for use in land mine detection applications. The array used in these experiments consists of 30 accelerometers in a 3×10 array spaced 3.429 cm apart in x and 10.287 cm in y. The measurements are interlaced along the y direction to synthetically generate a grid of measurement points with a spacing of $\Delta x = \Delta y = 3.429$ cm between measurement points in both x and y. By making many measurements, each at a different location on the surface, the acceleration of the entire scan region can be constructed synthetically. After the entire scan has been completed, a data array of acceleration information is available, $A(x_i, y_j, t_k)$, where
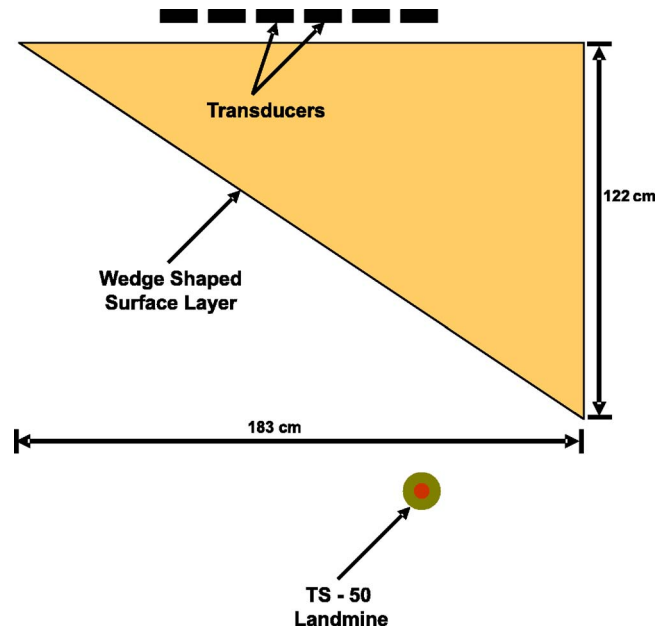
Fig. 2. (Color online) The experimental configuration for experiments using an asymmetric surface layer with a TS-50 land mine at the focusing point.

$$x_i = i\Delta x, \quad i = 0, 1, \ldots, \frac{X}{\Delta x}$$

$$y_j = j\Delta y, \quad j = 0, 1, \ldots, \frac{Y}{\Delta y}$$

$$t_k = k\Delta t, \quad k = 0, 1, \ldots, \frac{T}{\Delta t}, \tag{1}$$

and where $X = 202.311$ cm and $Y = 150.876$ cm are the dimensions of the scan region, and $T = 4.096$ s is the duration of time for which each measurement is recorded.

Time-reversal drive signals are created by first transmitting individual signals from each source in the array of sources and recording the response at the desired focus point. These signals are then time-reversed and retransmitted from the sources and the response is recorded over the entire scan region. The signal used to interrogate the soil is a swept frequency chirp signal that sweeps from 30 Hz to 2 kHz.[15] For display purposes, the results are presented as the response to a differentiated Gaussian pulse with a center frequency of 900 Hz. The process used to create the time-reversal drive signals is described in earlier work.[15,16]

2.2 Surface layer

An asymmetric surface layer is introduced into the experiment by embedding a plywood layer into the sand. The purpose of this layer is to alter the propagation direction of surface waves. The plywood layer is created using a wedge-shaped piece of 3/4 in. plywood cut into a 4 ft by 6 ft wedge. The shape of the surface layer is made asymmetrical with respect to the propagation direction of the surface waves as shown in Fig. 2.

The plywood wedge is tested in two configurations. The first configuration employed a single sheet plywood wedge that is buried flush to the surface of the sand (Case A) [Fig. 3(a)]. A second configuration uses two identical 4 ft by 6 ft by 3/4 in. plywood wedges. The upper sheet
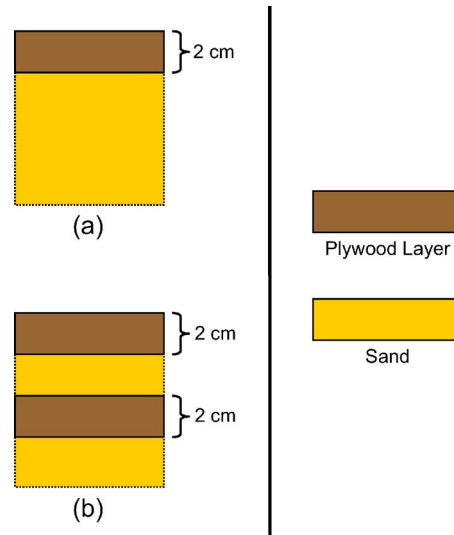
Fig. 3. (Color online) The *X-Z* plane cross section of the two surface layer configurations used in the experiments: (a) single sheet plywood surface layer, (b) multilayer plywood and sand surface layer.

of plywood was buried flush to the surface, with a thin layer (2 cm) of sand between the upper and lower sheets (Case B) [Fig. 3(b)]. In each configuration, a TS-50 land mine is buried at the focus location as indicated in Fig. 2.

3. Results: Focusing through an asymmetric surface layer

A comparison can be made of the performance of the three different excitation types; time-reversal focusing, time-delay focusing, and uniform excitation. Uniform excitation excites all sources with identical, in-phase signals. Time-delay focusing excites sources with signals that are time-delayed based on a constant velocity estimate for the Rayleigh wave propagation speed in the medium. If the estimate is accurate, the dominant Rayleigh wave from all the sources will arrive and focus coherently at a focus location at the same time. A comparison of these two excitation methods to time-reversal focusing is performed for both configurations of the asymmetric surface layer (Fig. 4).

The results are presented in Fig. 4 as pseudocolor graphs and animations of the magnitude of the vertical component of the particle acceleration at the surface. The pseudocolor scale used in the figures is a 40 dB logarithmic scale from white (0 dB) to black (−40 dB). In order to compare the different focusing methods, the results are normalized with respect to the energy contained in the excitation signals for each of the three excitation methods. An examination of time-domain movies and snapshots of wave propagation for each type of excitation is useful to visualize the effects of a surface layer on surface wave propagation.

3.1 Case A: Single sheet plywood surface layer

Because of the higher propagation speed in the plywood than in sand, the wedge shape of the plywood layer causes the surface-bound wave to turn in Fig. 4 (Case A). A second wave front travels under the surface layer, coming back to the surface of the sand on the other side of the plywood wedge. As can be seen in Fig.4 (Case A), this wave arrives at the land mine location later in time. As its speed is less affected by the plywood wedge on the surface, the change in its propagation direction is almost imperceptible.

While the direction change of the second wave front is small, the initial wave front is turned away from the land mine location. This initial wave contains a substantial portion of the excitation energy. Because of this, the maximum amplitude at the land mine location using
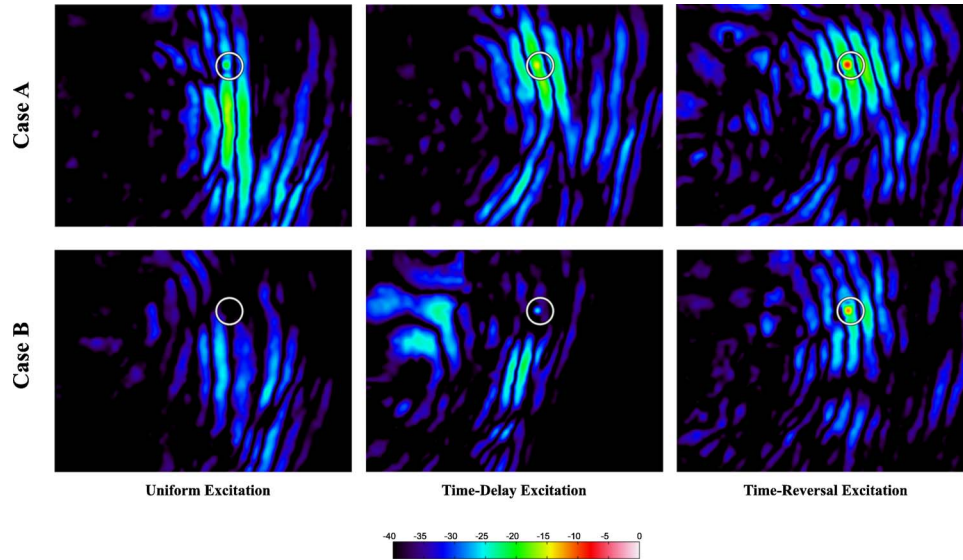
Fig. 4. (Color online) Time-domain animations of wave propagation through the single sheet plywood surface layer (Case A) and the multilayer plywood and sand surface layer (Case B). The white circle denotes the desired focus location, and the location of a buried TS-50 land mine. Images are on a 40 dB pseudocolor scale: 0 dB (white) to −40 dB (black). (Click on the image to view the animation.)

time-delay focusing is adversely affected. Time-reversal focusing shows a 7 dB improvement in peak excitation amplitude over time-delayed focusing and a peak signal level of approximately 13 dB over uniform excitation (Table 1).

3.2 Case B: Multilayer plywood and sand surface layer

The multilayer plywood and sand surface performs similarly to the single sheet plywood surface layer except that it turns the wave more effectively. Because the multilayer plywood and sand surface is more than twice the thickness of the single sheet plywood surface layer, more energy is captured in the surface layer, causing the relative amplitude of the turned wave to be larger in comparison to the second unturned wave. Most of the energy is turned away from the land mine location by the surface layer, as observed in Fig. 4 (Case B). Because of the discrepancy between the actual, inhomogeneous wave velocity, and the constant estimate used to calculate time delays, time-delay focusing concentrates energy in the wrong location. This effectively turns the wave away from the desired focus point. Time-reversal focusing shows an 8 dB improvement in peak excitation amplitude over time-delayed focusing and a peak signal level of approximately 18 dB over uniform excitation (Table 1).

Table 1. The peak amplitude (dB) and the background contrast (Δ dB) at the focus point for each excitation type and surface layer configuration.

Surface layer type	Uniform excitation (dB)	Time-delay excitation (dB)	Time-reversal excitation (dB)
Plywood single layer	−20	−14	−7
	Δ7	Δ5	Δ10
Plywood and sand multilayer	−28	−18	−10
	Δ5	Δ7	Δ10

4. Conclusions

The effectiveness of elastic wave time-reversal focusing was examined in the presence of an asymmetric surface layer. The surface layer changes the propagation velocity and direction of the elastic waves and can steer them away from the location of a buried land mine. The surface layer further complicated the wave field since waves could propagate under as well as through the surface layer. Even with these effects, time-reversal focusing was effective and performed significantly better than time-delay focusing. Table 1 summarizes the results.

[1] M. Fink, "Time-reversal of ultrasonic fields. I. Basic principles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 555–566 (1992).

[2] F. Wu, J.-L. Thomas, and M. Fink, "Time-reversal of ultrasonic fields. II. Experimental results," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 567–578 (1992).

[3] D. Casserau and M. Fink, "Time-reversal of ultrasonic fields. III. Theory of the closed time-reversal cavity," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 579–592 (1992).

[4] A. Derode, P. Roux, and M. Fink, "Robust acoustic time reversal with high-order multiple scattering," Phys. Rev. Lett. **75**, 4206–4209 (1995).

[5] M. Fink, C. Prada, F. Wu, and D. Cassereau, "Self focusing in inhomogeneous media with time reversal acoustic mirrors," Proc.-IEEE Ultrason. Symp. Vol. **5**, 681–686 (1989).

[6] P. Blomgren, G. Papanicolaou, and H. Zhao, "Super-resolution in time-reversal acoustics," J. Acoust. Soc. Am. **111**, 230–248 (2002).

[7] L. Borcea, G. Papanicolaou, and C. Tsogka, "Theory and applications of time reversal and interferometric imaging," Inverse Probl. Eng. **19**, S139–S164 (2003).

[8] R. Ing, M. Fink, and O. Casula, "Self-focusing Rayleigh wave using a time reversal mirror," Appl. Phys. Lett. **68**, 161–163 (1996).

[9] N. Chakroun, M. A. Fink, and F. Wu, "Time reversal processing in ultrasonic nondestructive testing," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 1087–1098 (1995).

[10] A. Sutin, J. TenCate, and P. Johnson, "Single-channel time reversal in elastic solids," J. Acoust. Soc. Am. **116**, 2779–2784 (2004).

[11] P. D. Norville and W. R. Scott, Jr., "An investigation of time reversal techniques in seismic land mine detection," J. Acoust. Soc. Am. **118**, 735–744 (2005).

[12] J. C. Santamarina, K. A. Klein, and M. A. F. N. Bracewell, *Soils and Waves* (Wiley, West Sussex, UK, 2001).

[13] W. R. Scott, Jr., J. S. Martin, and G. D. Larson, "Experimental model for a seismic land mine detection system," IEEE Trans. Geosci. Remote Sens. **39**, 1155–1164 (2001).

[14] J. S. Martin, G. D. Larson, and W. R. Scott, Jr., "An investigation of surface-contacting sensors for the seismic detection of buried land mines," J. Acoust. Soc. Am. **120**, 2676–2685 (2006).

[15] P. D. Norville, W. R. Scott, Jr., and G. D. Larson, "An investigation of time reversal techniques in seismic land mine detection," Proceedings of the SPIE Defense and Security Symposium, 2004, Vol. **5415**.

[16] P. D. Norville, "Time-reversal techniques in seismic detection of buried objects," Ph.D. thesis, Georgia Institute of Technology, 2007, Atlanta, GA (http://etd.gatech.edu/theses/available/etd-03292007–143303/).

Searching for best exemplars in multidimensional stimulus spaces

Eric Oglesbee and Kenneth de Jong
Department of Linguistics, Indiana University, Bloomington, Indiana 47405
eoglesbe@indiana.edu, kdejong@indiana.edu

Abstract: Examining phonetic categorization in multidimensional stimulus spaces poses a number of practical problems. The traditional method of forced identification becomes prohibitive when the number and size of stimulus dimensions becomes increasingly large. In response, Evans and Iverson [J. Acoust. Soc. Am. **115**, 352–361 (2004)] proposed an adaptive tracking algorithm for finding vowel best exemplars in a multidimensional space. This algorithm converged on best exemplars in a small number of trials; however, the search method was designed explicitly for vowel stimuli. In this paper, a more general multidimensional search algorithm is described, and results from simulations and experiments using the proposed algorithm are presented.
© 2007 Acoustical Society of America

1. Introduction

Virtually all speech contrasts involve multiple acoustic dimensions. Abramson and Lisker (1985) in a paper examining the effect of F0 contours on stop identification stated that "…research usually shows more than one cue to be pertinent to a distinction, although all such cues may not be equally important." Furthermore, Lisker (1986) discussed 16 acoustic properties that could plausibly be relevant to the *p-b* "voicing" distinction in the words *rapid* and *rabid*. Determining the relationship between these dimensions and the categorical contrast is an ongoing issue in perceptual and acoustic speech research; however, there are many practical barriers to examining a large number of acoustic properties in a single study.

A typical method for examining category locations is to generate a one-dimensional stimulus continuum, randomize the stimuli, and then present them multiple times to listeners for forced identification [e.g., Lisker and Abramson (1970); also Kitahara (2001) is an excellent example of using forced identification on multidimensional stimulus continua]. While this approach is feasible when the number of stimuli is small, it rapidly becomes unwieldy in multidimensional stimulus spaces. For example, assuming 30 judgments per stimulus and 2 s per response, it would take 16.7 h of nonstop data collection in order to work through a three-dimensional stimulus space containing ten stimulus steps per dimension. If a fourth ten-step dimension is added the amount of time balloons to 6.94 days, and a fifth ten-step dimension would push the time to 69.44 days.

An alternative approach for two-dimensional spaces was proposed by Johnson *et al.* (1993) where listeners navigated a two-dimensional grid on a computer screen in order to locate best exemplars. Each button in the grid corresponded to a single test stimulus. Calling this approach a "Method of Adjustment (MOA)" task, Harnsberger *et al.* (2001) used this strategy to study perceptual vowel spaces of cochlear implant users. While effective for two-dimensional, and even small three-dimensional spaces, from a practicality standpoint this approach does not scale up well to larger *n*-dimensional stimulus spaces.

Evans and Iverson (2004) proposed a novel method in an attempt to overcome these limitations. They used goodness ratings provided by listeners to steer an adaptive tracking algorithm through a four-dimensional space designed to find best exemplars of vowel categories. Although their algorithm found reliable estimates for best exemplars after only 30 trials per

vowel category, their search method took advantage of prior knowledge as to the likely locations of vowels in their stimulus space in order to achieve rapid convergence. In particular, their algorithm used the notion of schwa as the "neutral" vowel to initially direct the search for best exemplars along a diagonal passing through the expected location of the best exemplar and schwa in an F1-F2 plane. While this proved to be effective for quickly identifying best exemplars of vowels, the consequence is that the functioning of their algorithm is intertwined with properties of the stimulus space, and therefore not necessarily appropriate for other multidimensional spaces investigating other types of categories.

In this letter, a generalized *n*-dimensional search method is proposed [Algorithm for finding Multidimensional Best Exemplar Locations (AMBEL)], and results from simulations using random goodness judgments are presented.

2. AMBEL

A key feature of the Evans and Iverson (2004) approach that is incorporated here is the use of goodness ratings from listeners to direct the search for best exemplars in a multidimensional space. However, unlike Evans and Iverson (2004), AMBEL is designed to always use multiple iterations of the search process to achieve convergence. Section 2.1 describes a single iteration of the general AMBEL approach in a two-dimensional stimulus space. Section 2.2 contains a brief discussion of key features of the algorithm, as well as specific implementation details.

2.1 General procedure

In Fig. 1, two dimensions (D1 and D2) containing 13 stimuli each are represented. Each point in the grid corresponds to a single stimulus. Here, the search order for the dimensions is arbitrarily D1 followed by D2. In the first step, an initial point for the search is chosen and two comparison candidates are chosen varying in opposite directions from the initial point. Thus, a range of the probed space is chosen as a search vector. The search vector is constrained to varying D1 only, and by design does not span the entire length of D1. Second, the middle point of the search vector is played for the listener, and a goodness judgment is elicited using a slider bar that is part of a graphical interface. Prior to assigning the goodness judgment, listeners are allowed to play the stimulus as many times as they wish. Next, an end point is chosen at random, and the listener assigns a goodness judgment, followed by the opposite end point. The three goodness judgments obtained from the initial point and end points are then used to estimate the location of a "best derived point" in the search vector. This best derived point then serves as the initial point of the next search vector, which will probe D2, and the method for obtaining initial point and end point goodness judgments is repeated for D2. In order to achieve convergence, the process iterates through all of the dimensions multiple times. In the case of Fig. 1, the cross symbol in the lower right panel would become the new initial point for probing D1 if a second iteration was used. Although the above-presented example is given in two dimensions, the process scales upward to *n* dimensions.

2.2 Discussion of key features

AMBEL embodies six key features:

1. *Search vectors are constrained to a single dimension*. Searching a single dimension at a time removes the need for making *a priori* decisions concerning relationships between stimulus dimensions. This makes AMBEL suited for probing stimulus spaces in which relationships between dimensions are unknown.

2. *Best exemplars in each dimension are derived from goodness judgments taken from three points contained in the search vector (initial point, left end point, right end point)*. In principle, AMBEL allows for freedom of choice by the experimenter with regards to how the goodness judgments are used to derive the "best point" on each search vector. In our lab, goodness judgments were recorded on a scale ranging from 0.01 to 1.00. A nonzero lower bound was used to avoid division by zero when calculating the goodness ratio. In simulations and pilot experiments, the location of the best point in a search vector was determined by taking the ratio
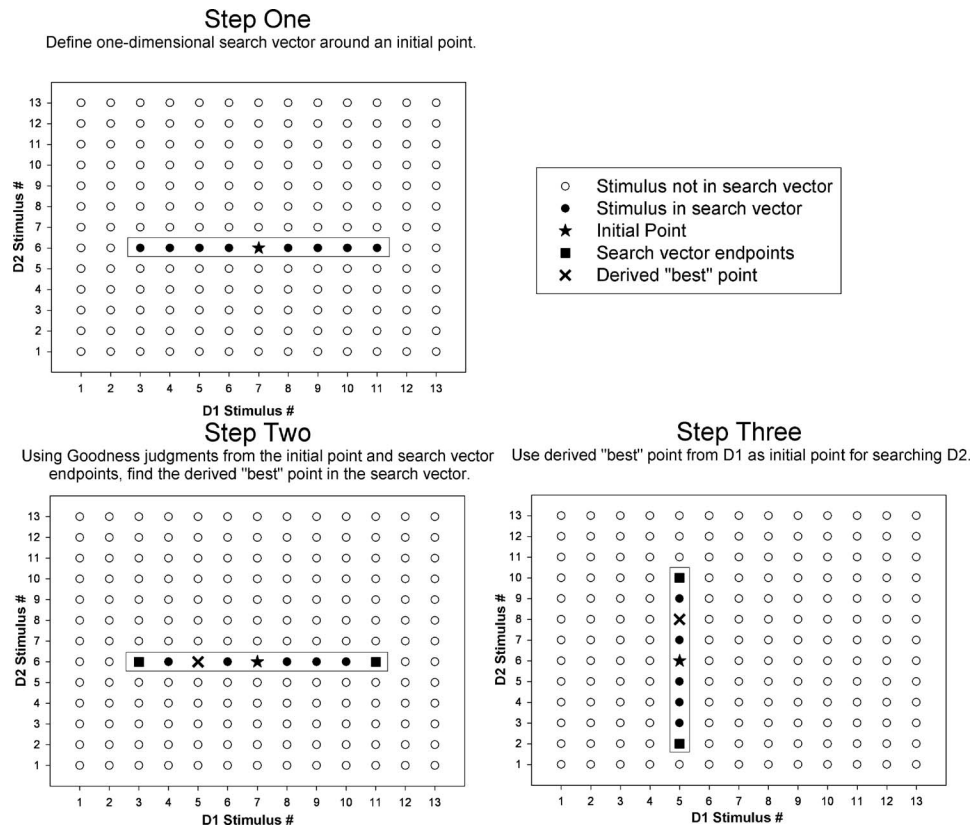
Fig. 1. Example of a single iteration of the search algorithm procedure in a two-dimensional space. Boxed in regions indicate current search vectors. If a second iteration were shown, the derived "best" point in the lower right panel (crosses) would serve as the initial point for searching D1.

of the two highest goodness judgments and modulating it by the decreasing exponential function in Fig. 2 (solid line).

The location of the best point was expressed as a proportion of the distance between the two stimuli with the highest goodness judgments as measured from the highest rated stimulus. Modulating the goodness judgment ratio by the decreasing exponential in Fig. 2 (solid line) has the advantage of moving the choice of the "best" exemplar toward the stimulus with the highest goodness judgment further than would be true with a simple ratio (see the dotted curve in Fig. 2). For example, if a simple ratio was being used and the goodness ratings were in a 2:1 relationship, the location of the derived best point would be 33% of the distance from the highest rated point to the second highest rated point. However, using the decreasing exponential function in Fig. 2 to modulate the ratio, the derived best point is located at 10% of the distance. This weighting biases the selection of the derived point toward a location that the listener has already identified as being the best of the three stimuli that have been presented, and it also allows movement away from the center of the space.

3. *The estimate used for eliciting goodness judgments is retained between stimulus presentations.* The slider bar is not reset at the beginning of each stimulus presentation. Not resetting the slider bar between goodness judgments allows for consecutively played stimuli to be judged relative to one another. An additional consequence of this is that the search process generates two additional pieces of data that are quite valuable. First, since the search vector end points are always played consecutively within a dimension, the listener is directly comparing the end points. Movement in the slider bar after presentation of the second end point, then
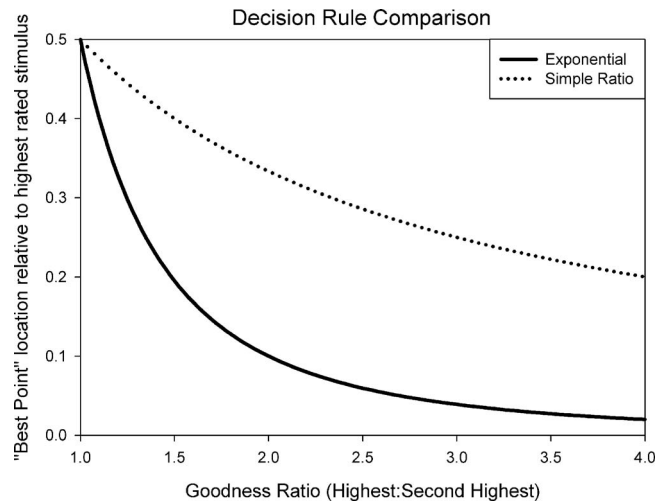
Fig. 2. Decreasing exponential function used to modulate goodness ratios. The function is $X(r)=0.5*r^{-2.3219}$, where $r=g_1/g_2$, g_1 is the highest goodness judgment, and g_2 is the second-highest goodness judgment. The goodness ratio "r" is shown on the horizontal axis, and the distance of the "best" exemplar from the stimulus associated with goodness judgment g_1 is given on the vertical axis as a proportion of the distance between the stimuli associated with g_1 and g_2.

indicates the listener's preference for changes in D1. These search vector end point preferences provide useful data concerning listeners' tendency of responding to differences in a particular dimension. Second, the goodness of the derived best point can be directly compared to the stimulus that immediately preceded it (one of the search vector end points). If the algorithm for choosing derived best points is consistently steering the search into regions containing better exemplars, the derived best point will have a goodness value that is the same or better than the point that preceded it. This can be used as a post hoc test of the search algorithm's performance.

4. *Only a subset of a stimulus dimension is searched at a given time*. A search vector is large enough to choose acoustically very different stimuli, causing listeners to use a large range of the slider bar when presented initial point and end point stimuli, but small enough to allow a large set of possible stimuli to be played during the course of multiple iterations. In our simulations and experiments, search vectors were constrained to a maximum length of 2/3 of a stimulus dimension. In cases where the initial point was near an edge of the space, the side of the search vector closest to the edge of the space was truncated.

5. *All dimensions are probed before a dimension is probed again*. Every dimension is "tuned" before a particular dimension is tested again. In principle, a fixed or random ordering of dimensions within an iteration can be used.

6. *Multiple iterations of the search process are used to achieve convergence*. The primary principle behind the proposed search method is that incremental progress made in each individual dimension aids global convergence on a best exemplar. By iterating through all of the dimensions multiple times, listeners are given multiple opportunities to steer the algorithm into the proper portion of the space. Although more efficient approaches are possible, the robustness of the proposed method, as well as the ability to identify effects of individual dimensions, seems to offset these inefficiencies.

3. Results from simulations and pilot experiments

Oglesbee and de Jong (2006) present an analysis of the properties of the proposed search algorithm obtained from simulations using random goodness judgments. They examined how the probability of a particular stimulus being the final destination point of the algorithm is affected by the initial starting location of the search, number of iterations, and number of stimuli in each
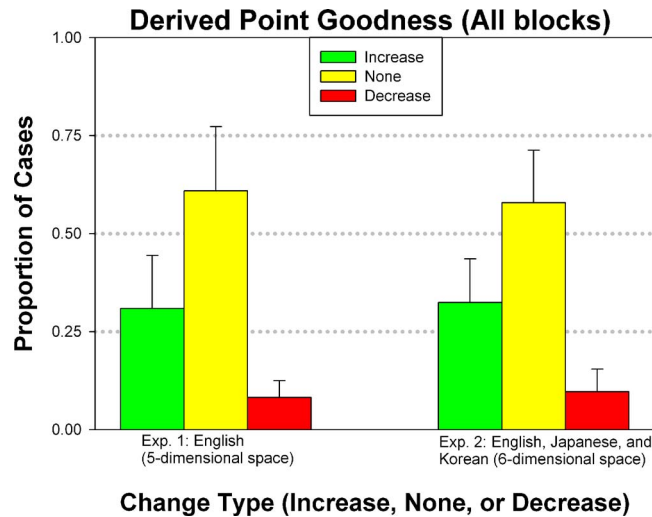
Fig. 3. (Color online) Proportion of cases where the derived point goodness was the better, the same, or worse than the stimulus preceding it. For each experiment, data were pooled across all listeners and categories tested. Error bars indicate 1 s.d.

dimension. Results showed that if the number of iterations was sufficiently large, each stimulus in a dimension has a nearly equal chance of being selected, although dimension end points were mildly less likely to be selected than other points. This is a positive feature since it indicates that the algorithm does not trap listeners against the edges of the stimulus space. If the initial point of the search is located near the center of a dimension, convergence to the long-term shape of the probability distribution is very rapid (approximately three iterations).

  A second way to evaluate the effectiveness of the algorithm is to look at the derived best point goodness data arising from experiments using human listeners. Figure 3 contains histograms depicting the proportion of cases where the derived best point had a goodness judgment that was better, the same, or worse as compared to the stimulus immediately preceding it. If the algorithm selects better stimuli based on listener provided goodness judgments, then the derived best point should have a goodness rating that is either the same or better than the previously presented stimulus. Since the slider bar is not reset between stimulus presentations, this is a very accurate measure of the relative goodness of the derived best point with respect to the stimulus preceding it. The data in Fig. 3 come from two experiments using AMBEL; experiments 1 [reported in Oglesbee and de Jong (2006)] and 2 examined word initial labial stops in five- and six-dimensional stimulus spaces, respectively. Experiment 1 was restricted to native speakers of English (5 listeners), whereas experiment 2 probed English (3 listeners), Japanese (3 listeners), and Korean (3 listeners).

  In both experiments the derived best point goodness was the same or better than the previously presented stimulus over 90% of the time. This gross measure of algorithm performance indicates that AMBEL does in fact steer listeners into regions containing better exemplars.

  Although the search process terminates at a specific point in a multidimensional stimulus space, this is not the only useful type of data generated by AMBEL. A discussion of additional types of data generated by the procedure can be found in Oglesbee and de Jong (2006).

4. Conclusions

4.1 Limitations

It is important to note that AMBEL is not a "magic bullet" for studying categories in multidimensional stimulus spaces. Rather, it is a first approximation tool that should be used to aid in

the development of more sensitive lower dimensional studies. Also, the goodness judgments elicited in the search process are meaningless in and of themselves. Since listeners are constantly recalibrating their use of the slider bar, the numerical values of the goodness judgments can only be used in a very restricted fashion. Namely, the goodness ratings of adjacently presented stimuli are the only ones that can be directly compared. Goodness ratings from near-adjacent stimuli should be compared with great caution, and ratings from different iterations or blocks should never be compared. (The only exception to this is that the last stimulus of an iteration can be compared to the first stimulus of the following iteration).

4.2 Generalizability and usage

One drawback of Evans and Iverson's (2004) algorithm is that their search method is specific to the properties of their stimulus space, and prior estimates of general category locations in that space. In contrast to this, the functioning of AMBEL is independent of the stimulus space being searched. It allows for searches through n-dimensional acoustic, and even multimodal, stimulus spaces. The only in principle constraint on the application of AMBEL is that it must be possible to elicit gradient goodness judgments with regards to the categories being examined.

AMBEL is a potentially powerful tool for examining categories cross-linguistically. Listeners from different language backgrounds can be tested on the same stimulus space, and given the constancy of the stimulus space and search algorithm, differences in performance can be attributed to subject characteristics (i.e., language background).

References and links

Abramson, A., and Lisker, L. (**1985**). "Relative power of cues: F0 shift versus voice timing," in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, edited by V. Fromkin (Academic, Orlando, FL), pp. 25–33.

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

Harnsberger, J. D, Svirsky, M. A., Kaiser, A. R., Pisoni, D. B., Wright, R., and Meyer, T. A. (**2001**). "Perceptual 'vowel spaces' of cochlear implant users: Implications for the study of auditory adaptation to spectral shift," J. Acoust. Soc. Am. **109**, 2135–2145.

Johnson, K., Flemming, E., and Wright, R. (**1993**). "The hyperspace effect: Phonetic targets are hyperarticulated," Language **69**, 505–528.

Kitahara, M. (**2001**). "Category structure and function of pitch accent in Tokyo Japanese," Doctoral dissertation, Indiana University.

Lisker, L. (**1986**). "Voicing in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees," Lang Speech **29**, 3–11.

Lisker, L., and Abramson, A. (**1970**). "The voicing dimension: Some experiments in comparative phonetics," in *Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967* (Academia, Prague), pp. 563–567.

Oglesbee, E., and de Jong, K. (**2006**). "Locating perceptual category centers in multidimensional stimulus spaces," IULC Working Papers Online, Vol. **6**., (https://www.indiana.edu/~iulcwp).

A comparison of vocal tract perturbation patterns based on statistical and acoustic considerations

Brad H. Story

Speech Acoustics Laboratory, Department of Speech, Language, and Hearing Sciences,
University of Arizona, Tucson, Arizona 85721
bstory@u.arizona.edu

Abstract: The purpose of this study was to investigate the relation between vocal tract deformation patterns obtained from statistical analyses of a set of area functions representative of a vowel repertoire, and the acoustic properties of a neutral vocal tract shape. Acoustic sensitivity functions were calculated for a mean area function based on seven different speakers. Specific linear combinations of the sensitivity functions corresponding to the first two formant frequencies were shown to possess essentially the same amplitude variation along the vocal tract length as the statistically derived deformation patterns reported in previous studies.

1. Introduction

Statistical analyses of collections of tongue configurations or complete vocal tract shapes (i.e., area functions) have revealed that a small number of canonical deformation patterns (variously referred to as factors, components, basis functions, or modes) can explain most of the variation in vocal tract shape during vowel production (Harshman *et al.*, 1977; Shirai and Honda, 1977; Jackson, 1988; Johnson *et al.*, 1993; Nix *et al.*, 1996; Story and Titze, 1998; Hoole, 1999; Zheng *et al.*, 2003; Iskarous, 2005; Story, 2005; Mokhtari *et al.*, 2007). These deformation patterns tend to exhibit similarities in shape across speakers and are related to specific formant frequency patterns when superimposed on a mean or neutral vocal tract shape.

As an example, mode shapes (deformation patterns) and mean area functions determined with principal component analysis are shown by the dotted lines in Fig. 1 for the one speaker of Story and Titze (1998) and the six speakers of Story (2005). The vocal tract length has been normalized so they can be easily compared. The thick line in each plot indicates the mean shape calculated across the seven speakers and is shown to summarize the general shape. The two modes shown accounted for at least 85% of the total variance in each speaker's collection vowel area functions. Mathematically, an arbitrary area function for a particular speaker can be represented as

$$V(x) = \frac{\pi}{4}[\Omega(x) + q_1\phi_1(x) + q_2\phi_2(x)]^2,$$
(1)

where x is the distance from the glottis, $\Omega(x)$ is the mean diameter function, $\phi_1(x)$ and $\phi_2(x)$ are the modes, and q_1 and q_2 are the weighting coefficients.[1] When used in Eq. (1) with a positive weighting coefficient, the first mode ϕ_1 for any of the speakers would have the spatial effect of expanding the front portion of the vocal tract while constricting the back, whereas a negative coefficient would have the opposite effect. Although there is variation among the speakers, in any of the cases a positively weighted second mode ϕ_2 would impose expansions in the lip and midtract regions, and constrictions posterior to the lips [between 0.65 and 0.85 in Fig. 1(b)] and just above the glottis. A negative coefficient would again create the opposite spatial effect.
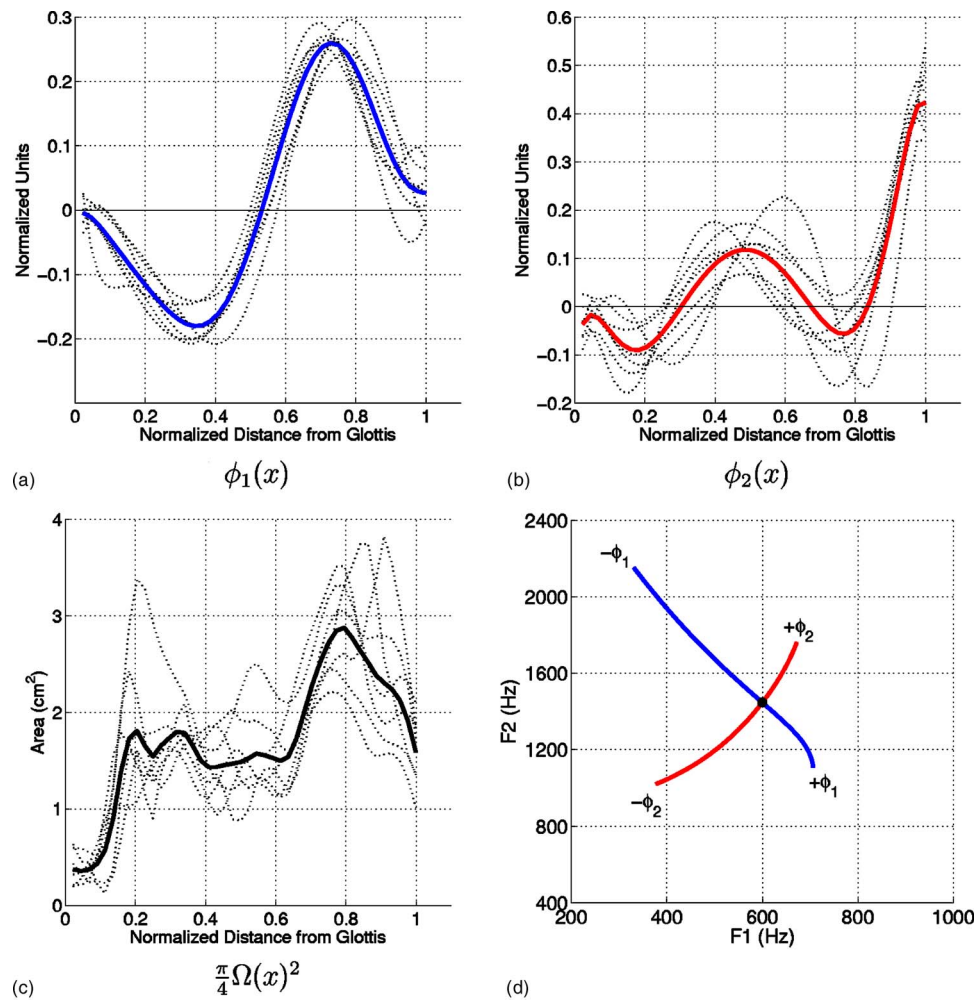
Fig. 1. (Color online) Mode shapes and mean area functions for seven speakers (dotted lines) from Story and Titze (1998) and Story (2005), along with the [F1,F2] plot produced by each mode isolation. The vocal tract lengths have been normalized to 1.0 so they can be overlaid for comparison purposes. The thick lines indicate the mean of the given function in each plot. (a) First mode ϕ_1, (b) second mode ϕ_2, (c) mean area function (MAF) $(\pi/4)\Omega^2$, and (d) [F1,F2] trajectories produced by independently superimposing ϕ_1 and ϕ_2 on the MAF.

To demonstrate the effects of each mode on the first two formant frequencies (F1 and F2), the mean ϕ_1 and ϕ_2 shown in Figs. 1(a) and 1(b) were superimposed on the mean vocal tract shape [Fig. 1(c)] according to Eq. (1). The scaling coefficients were incrementally varied as $q_1=[-4.5, 4.0]$ while $q_2=0$ and again as $q_2=[-2.5, 2.5]$ while $q_1=0$. Formant frequencies calculated for each of the series of area functions generated along the respective coefficient continua are plotted in the [F1,F2] space shown in Fig. 1(d). Relative to the neutral position denoted by the solid dot at $[600, 1450]$ Hz, a negatively scaled $-\phi_1$ decreases F1 while increasing F2, whereas a positive scaling increases F1 and decreases F2. For ϕ_2, a negative scaling coefficient will cause both F1 and F2 to decrease, whereas a positive scaling will have the opposite effect. The end points of each [F1,F2] trajectory roughly correspond to the vowels [i æ a u] in respective clockwise order, beginning in the upper left-hand corner of the plot. Although this demonstration is somewhat artificial because it is based on average modes across speakers, similar effects for F1 and F2 have been previously reported for all seven speakers. Various other

studies have also reported principal component analyses of area functions that resulted in modes shapes similar to those in Figs. 1(a) and 1(b) (Meyer *et al.*, 1989; Yehia *et al.*, 1996; Mokhtari *et al.*, 2007).

Although the modes are statistical constructs that describe a specific set of vocal tract data, their similarity across speakers suggests that they could represent some kind of generalized vocal tract shaping patterns that are produced and scaled by the speech motor system during vowel production. But why is it that these particular mode shapes emerge from the analyses of vocal tract data? It is possible that they are an artifact of the type of analysis performed. For instance, if a covariance matrix generated from a set of vocal tract shapes takes on Toeplitz form, the eigen-vectors (e.g., modes, components, basis functions) of that matrix will be sinusoidal (cf. Jolliffe, 2004) and could perhaps resemble the modes shown in Fig. 1. In such a case, the modes could be expected to reconstruct the original data with small error, but may not be related to anything specifically articulatory or acoustic. The systematic relation of the modes to the first two formant frequencies, however, suggests that their shapes emerge in order to exploit the acoustic properties of the vocal tract itself. In this view, the particular variation of each mode along the length of the vocal tract should reflect some representation of the pressure and volume velocity distribution that exists within the vocal tract at the resonance (formant) frequencies, and would predict how those frequencies should change when each mode is superimposed on a given vocal tract shape.

The concept that global changes in vocal tract shape during speech could be explained by the acoustic properties of a uniform tube was established by Schroeder (1967) and Mermelstein (1967). Using considerations of potential and kinetic energy densities, both showed that a uniform tube (i.e., an area function with constant area), of length comparable to a human vocal tract, could be systematically perturbed with a superposition of a series of sinusoids to produce area functions that supported a particular set of formant frequencies. Similar acoustic theory was later used by Fant and Pauli (1975) to predict the direction of formant frequency change when small perturbations were applied to a specific area function. They calculated acoustic "sensitivity functions," which quantify the difference between potential and kinetic energy at each formant frequency as a function of the distance along the vocal tract. Thus, these functions can be used to indicate which parts of the area function should be expanded or contracted in order to move a formant frequency toward a desired value. Mrayati *et al.* (1988) made extensive use of sensitivity functions to develop a model, called the distinctive region model (DRM), in which a uniform tube representation of the vocal tract could be divided into a small number of regions (four to eight depending on the number of formants to be controlled), each of which could impose a "distinctive" change in the formant frequencies when expanded or contracted. Although the DRM was criticized for being too simplistic as a comprehensive model of speech production (cf. Boe and Perrier, 1990), it clearly demonstrated that formant frequencies could be efficiently moved when tube shape changes were roughly aligned with the distinctive regions. More recently, Carré (2004) has reported similar results when the vocal tract tube shape is perturbed with scaled versions of the sensitivity functions themselves, rather than the discrete regions of the DRM. Carré (2004) emphasizes that the shape changes imposed on a uniform tube that produce speech-like formant patterns are based purely on the acoustic properties of that tube, i.e., other than approximate length, no *a priori* knowledge of the human vocal tract was assumed. If such theoretically based shape changes are truly indicative of those produced during speech, then they should be well correlated with vocal tract shaping patterns derived from human articulatory data.

Using the factors derived from tongue configuration data reported by Harshman *et al.* (1977), Fitch *et al.* (2003) developed a sinusoidal model of vocal tract shape. That is, two sinusoids were used to approximate the effect of the factors on the vocal tract area function. They noted a correspondence between the shapes of linear combinations of the sinusoidal components in the model and the variation of the sensitivity functions for F1 and F2. The implication was this particular representation of articulatory patterns did indeed exploit the acoustic sensitivity of the vocal tract. At nearly the same time, Ru *et al.* (2003) reported a similar sinusoidal

model of the area function that was also based on the Harshman *et al.* factors. Whereas the goals of this study were different than Fitch *et al.*, an equivalence between the acoustic characteristics of the vocal tract and the sinusoidal components was noted.

Although the area function-based modes share some similarities with the factors reported by Harshman *et al.* (1977) and others, they are not identical, and hence their relation to sensitivity functions is expected to be somewhat different. The purpose of this letter is to demonstrate that each of the two modes derived from principal component analyses of area functions (e.g., Fig. 1) corresponds to specific linear combinations of acoustic sensitivity functions. It will be shown that such a correspondence provides some explanation as to why these common mode shapes are observed across speakers for vowel production.

2. Sensitivity functions

As a demonstration case, sensitivity functions were calculated for the "mean of the mean area functions" shown by the dark line in Fig. 1(c). This idealized mean area function, henceforth referred to as "MAF," contains features typical of the mean area functions that have been calculated for various speakers. For example, the initial 10% (0.1) of the normalized tract length tends to coincide with small cross-sectional areas [approximately 0.4 cm^2 in Fig. 1(c)] and is sometimes referred to as the epilaryngeal space. The pharyngeal portion extending from about 0.2 to 0.6 (of normalized length) is fairly constant with an area of about 1.5 cm^2, whereas in the oral cavity there is a moderate expansion.

For computational purposes MAF was represented by two vectors, $a(i)$ and $l(i)$, which are the cross-sectional areas and lengths, respectively, of each of 44 sections ($i = [1, \ldots, 44]$) extending along the vocal tract from glottis to lips. Because the sensitivity function calculation requires an actual vocal tract length (rather than a normalized length), the 44 sections of the length vector were each assigned a value of $l(i) = 0.4$ cm. This choice generates a total tract length of 17.6 cm, which is typical of an adult male speaker. Although it may seem inappropriate to impose a male tract length on an area function that is based on both male and female speakers, the amplitude variation of the sensitivity functions along the tract length is unaffected by uniform length scaling. Hence, the choice of section length is arbitrary and the normalized length axis will be maintained for subsequent plots.

The method for calculation of sensitivity functions for this study was identical to that described in Story (2006), in which pressures, flows, frequency response functions, and kinetic and potential energies were determined with a transmission-line type model of the vocal tract (e.g., Sondhi and Schroeter, 1987; Story *et al.*, 2000) that included energy losses due to yielding walls, viscosity, heat conduction, and acoustic radiation at the lips. The sensitivity of a specific formant frequency to a change in cross-sectional area can be defined as the difference between the kinetic energy and potential energy as a function of distance from the glottis, divided by the total energy in the system (Fant and Pauli, 1975). A set of sensitivity functions $S_n(i)$ can be determined for the resonance frequencies (formants), F_n, of any given area function $a(i)$, where n is the formant number. In theory, sensitivity functions, $S_n(i)$, can be used to compute the change in a particular formant frequency (F_n) due to perturbation of the area function (Δa) with the relation

$$\frac{\Delta F_n}{F_n} = \sum_{i=1}^{N_s} S_n(i) \frac{\Delta a(i)}{a(i)}, \qquad (2)$$

where N_s is the number of sections comprising the area function. When the sensitivity function $S_n(i)$ is positively valued and an area perturbation is also positive (i.e., area is increased), the change in formant frequency will be upward. If the area change is negative (area decreased) the formant frequency will decrease. When the sensitivity function is negatively valued, the opposite effect occurs for positive or negative area perturbations, respectively.

The sensitivity functions, S_1 and S_2, calculated for MAF are shown in Fig. 2(a). Each line extends along the distance from the glottis to lips and indicates the relative sensitivity of the
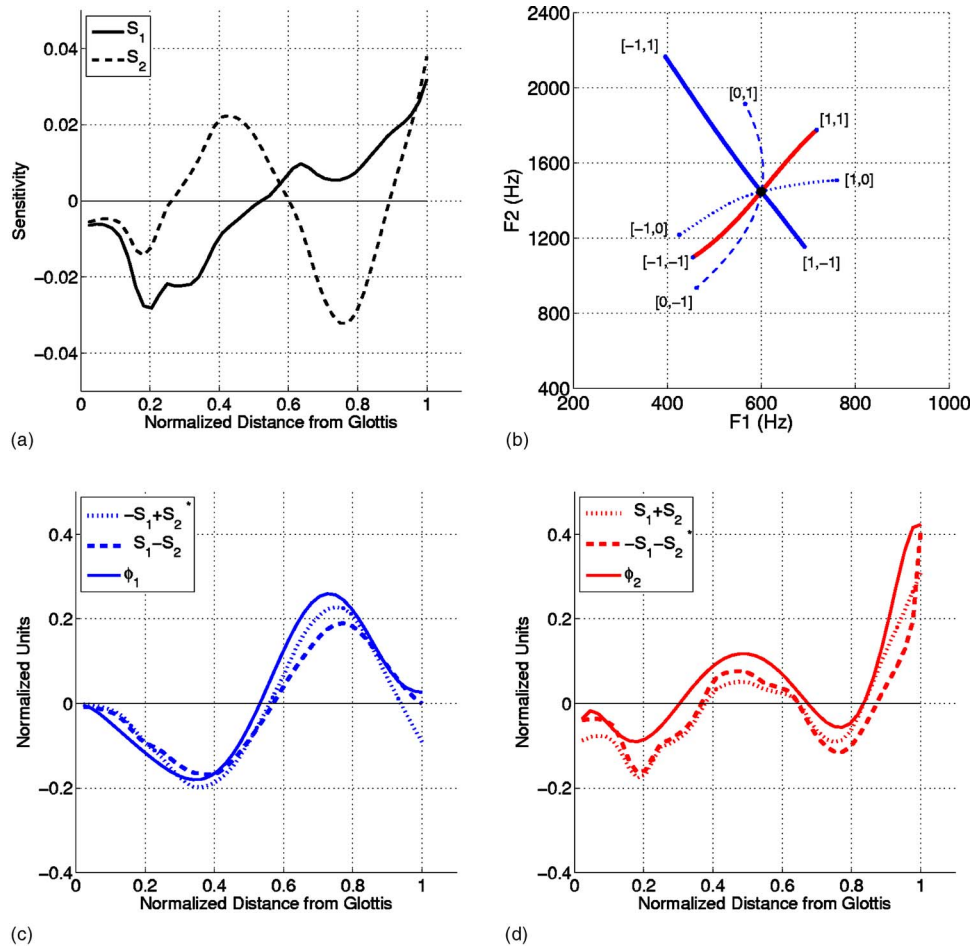
Fig. 2. (Color online) (a) Sensitivity functions calculated for the first two resonances of MAF [dark line in Fig. 1(c)], (b) [F1,F2] trajectories produced by Eq. (4) with eight different settings of the z_1 and z_2 coefficients, (c) comparison of ϕ_1 (solid line) to the mean $(-S_1+S_2)$ (dotted line) and mean (S_1-S_2) (dashed line), and (d) comparison of ϕ_2 (solid line) to the mean (S_1+S_2) (dotted line) and mean $(-S_1-S_2)$ (dashed line). In (c) and (d), the sensitivity function combinations have been linearly scaled so that they have roughly the same amplitude as the modes and those with an asterisk have been flipped in polarity purely for visual comparison purposes.

first and second formants (F1 and F2) to a small perturbation of the area function $[\Delta a(i)]$. Using S_1 in Fig. 2(a) and Eq. (2) as a guide, it is observed that F1 could be increased by expanding the area in the front half (0.5–1.0 of the normalized tract length) of the vocal tract. F1 could also be increased by constricting the regions between the glottis and halfway to the lip termination. 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 0.25–0.6 and 0.9–1.0 of the normalized length, and constricting the regions of the area function that extend from 0 to 0.25 and 0.6 to 0.9; lowering F2 would require the opposite changes in area.

3. Comparison of sensitivity function perturbations and vocal tract modes

The predictions afforded by the shapes of the sensitivity functions in Fig. 2 suggest that formant frequencies could be controlled and positioned (in the acoustic domain) by perturbing the original vocal tract shape with replicas of the sensitivity functions themselves (Carré, 2004; Story,

2006). For example, direct superposition of S_1 on MAF would raise F1, whereas its opposite, $-S_1$, would lower it. Similarly, F2 could be controlled with a superposition of a scaled S_2 replica, where $+S_2$ would increase F2 and $-S_2$ would decrease it.

It follows that simultaneous modification of F1 and F2 could be realized with a superposition of both S_1 and S_2. For example, when appropriately scaled to affect cross-sectional area and superimposed on the original area function, the combination $(S_1 - S_2)$ would be expected to alter the vocal tract shape such that F1 increases and F2 decreases. An arbitrary area function can be described mathematically as

$$a_{\text{new}}(i) = a_0(i) + [z_1 S_1(i) + z_2 S_2(i)], \quad i = [1, N_{\text{areas}}], \tag{3}$$

where $a_0(i)$ is the area function on which $S_1(i)$ and $S_2(i)$ are based, and $a_{\text{new}}(i)$ is a new area function generated by the superposition of the linear combination. The z_1 and z_2 are scaling coefficients that, for the above-presented example, would be equal to 1 and -1, respectively. Because the sensitivity functions are dependent on a particular vocal tract shape, the prediction of formant frequency change is limited to small area changes. Thus, generating vocal tract deformations comparable to those during vowel production requires an iterative perturbation process in which $a_{\text{new}}(i)$ in Eq. (3) replaces $a_0(i)$ in subsequent iterations. Equation (2) can be recast as

$$a_{k+1}(i) = a_k(i) + [z_1 S_{1_k}(i) + z_2 S_{2_k}(i)], \quad i = 1[1, N_{\text{areas}}], \quad k = [0, N_{\text{iter}}], \tag{4}$$

where the $a_k(i)$'s and S_{n_k}'s are area vectors and sensitivity functions, respectively, at successive iterations, and k is the iteration index, which ranges from 0 (to denote the initial area function) to the number of desired iterations (N_{iter}).

Using Eq. (4) with MAF as the initial vocal tract shape $a_0(i)$, a series of area vectors was generated for each of eight different settings of the z_1 and z_2 coefficients. These consisted of all combinations of $z_1 = [-1, 0, 1]$ and $z_2 = [-1, 0, 1]$ (except for [0,0]) where, in each case, area vectors were generated for a maximum of 50 iterations or until the minimum area was equal to 0.2 cm^2. The [F1,F2] formant trajectories corresponding to each coefficient setting are shown in Fig. 2(b) as projecting outward from a central point determined by the formant frequencies of MAF. The dotted and dashed lines indicate the effect of a perturbation based on only the positive and negative polarities S_1 or S_2, respectively, i.e., either z_1 or z_2 was zero. The S_1 trajectory (dotted) primarily traverses the F1 dimension but does curve upward in F2 on the [1,0] side and downward on the [−1,0] side. In contrast, the S_2 trajectory (dashed) moves mostly along the F2 dimension, but curves downward (in frequency) in the F1 dimension at both ends. The other two trajectories, shown as solid lines in Fig. 2(b), are the result of the four linear combinations of both S_1 and S_2 for which z_1 and z_2 were equal to either 1 or −1. The trajectory corresponding to [−1,1] and [1,−1] is nearly linear and extends from a region of low F1 and high F2 to a region of high F1 and low F2, the end points of which would roughly correspond to the vowels /i/ and /a/ or /ɔ/. A nearly opposite change in formant frequencies is traced out by the trajectory corresponding to [−1,−1] where F1 and F2 are both low, and [1,1] where both F1 and F2 are high. This trajectory is also fairly linear and the end points would approximately correspond to the vowels /u/ or /o/ at the lower left-hand corner and /ae/ at the upper right.

These latter two trajectories nearly replicate the paths through the [F1, F2] space produced by the ϕ_1 and ϕ_2 modes presented previously in Fig. 1(d). Since the initial area function (MAF) was the same for both the mode-based and sensitivity-based perturbations, it should follow that the perturbation shapes of either method should be similar. Shown in Fig. 2(c) are the means of the linear combinations $(S_1 - S_2)$ and $(-S_1 + S_2)$ plotted along with the ϕ_1 mode from Fig. 1(a). For ease of visual comparison, the amplitudes of each function have been similarly scaled and the polarity of the $(-S_1 + S_2)$ perturbation has been intentionally flipped. With the exception of a zero crossing offset near the middle of the vocal tract and a slight difference in amplitude at the lips, the two sensitivity-based perturbations appear to be nearly identical to ϕ_1. To quantify the similarity, a correlation coefficient (R) was calculated for each sensitivity-

based perturbation relative to ϕ_1, and resulted in both being equal to $R=0.97$. A similar plot is shown in Fig. 2(d), where the ϕ_2 mode is compared to the means of the linear combinations ($-S_1-S_2$) and (S_1+S_2). Although the similarity is perhaps not as visually distinct as in the previous case, the correlation coefficients for the two sensitivity-based perturbations relative to ϕ_2 are 0.92 and 0.97, respectively. It can also be observed that each of the three perturbation functions would generate expansions and constrictions in nearly the same locations along the tract length. The main differences consist of slight offsets of the zero crossing locations and a greater amplitude of the negative portions of ($-S_1-S_2$) and (S_1+S_2).

4. Discussion

The results of this study suggest that the shape of the statistically based area function modes can be approximately related to acoustic sensitivity functions with the following equivalences:

$$+ \phi_1 \equiv (+ S_1 - S_2) \quad [F1\uparrow F2\downarrow],$$

$$- \phi_1 \equiv (- S_1 + S_2) \quad [F1\downarrow F2\uparrow],$$

$$+ \phi_2 \equiv (+ S_1 + S_2) \quad [F1\uparrow F2\uparrow],$$

$$- \phi_2 \equiv (- S_1 - S_2) \quad [F1\downarrow F2\downarrow], \tag{5}$$

where the directions of formant frequency changes due to a superposition of either a mode or sensitivity function combination are shown in the right column. The implication is that the vocal tract shaping patterns (i.e., ϕ_1 and ϕ_2) determined through statistical analyses of sets of area functions represent essentially the same spatial variation along the length of the vocal tract as do these linear combinations of the sensitivity functions. In either case, it is noteworthy that the perturbations move F1 and F2 from a neutral location toward the *corners* of the vowel space that approximately correspond to [i æ a u]. Carré (2004) reported similar trajectories and noted that shape changes based on sensitivity function combinations could achieve a given acoustic contrast more effectively than those based on a single sensitivity function. That is, vocal tract perturbation patterns that create simultaneous changes to both F1 and F2 provide an efficient means by which to navigate the vowel space. Thus, the equivalence of the ϕ_1 and ϕ_2 modes with particular combinations of S_1 and S_2 would seem to emerge because these shapes allow for efficient traversal to the extreme regions of the acoustic vowel space.

Acknowledgment

References and links

[1]Since the principal components analysis was performed on the equivalent diameters of each, cross-sectional area within the area function sets, the squaring operation and scaling by $\pi/4$ are needed to convert diameter to area [see Story (2005) for details].

Boë, L. J., and Perrier, P. (**1990**). "Comments on 'Distinctive regions and modes: A new theory of speech production' by M. Mrayati, R. Carré and B. Guérin," Speech Commun. **9**, 217–230.

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

Fant, G., and Pauli, S. (**1974**). "Spatial characteristics of vocal tract resonance modes," in Proceedings of the Speech Communication Seminar 74, Stockholm, Sweden, 1–3 August, pp. 121–132.

Fitch, H. L., Kupin, J. J., Kessler, I. J., and Delucia, J. (**2003**). "Relating articulation and acoustics through a sinusoidal description of vocal tract shape," Speech Commun. **39**, 243–268.

Harshman, R., Ladefoged, P., and Goldstein, L. (**1977**). "Factor analysis of tongue shapes," J. Acoust. Soc. Am. **62**, 693–707.

Hoole, P. (**1999**). "On the lingual organization of the German vowel system," J. Acoust. Soc. Am. **106**, 1020–1032.

Iskarous, K. (**2005**). "Patterns of tongue movement," J. Phonetics **33**, 363–381.

Jackson, M. T. T. (**1988**). "Analysis of tongue positions: Language-specific and cross-linguistic models," J. Acoust. Soc. Am. **84**, 124–143.

Johnson, K., Ladefoged, P., and Lindau, M. (**1993**). "Individual differences in vowel production," J. Acoust. Soc. Am. **94**, 701–714.

Jolliffe, I. T. (**2004**). *Principal Component Analysis*, 2nd ed. (Springer, New York).

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

Meyer, P., Wilhelms, R., and Strube, H. W. (**1989**). "A quasiarticulatory speech synthesizer for German language running in real time," J. Acoust. Soc. Am. **86**, 523–539.

Mokhtari, P., Kitamura, T., Takemoto, H., and Honda, K. (**2007**). "Principal components of vocal tract area functions and inversion of vowels by linear regression of cepstrum coefficients," J. Phonetics **35**, 20–39.

Mrayati, M., Carre, R., and Guerin, B. (**1988**). "Distinctive regions and modes: A new theory of speech production," Speech Commun. **7**, 257–286.

Nix, D. A., Papcun, G., Hogden, J., and Zlokarnik, I. (**1996**). "Two cross-linguistic factors underlying tongue shapes for vowels," J. Acoust. Soc. Am. **99**, 3707–3717.

Ru, P., Chi, T., and Shamma, S. (**2003**). "The synergy between speech production and perception," J. Acoust. Soc. Am. **113**, 498–515.

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

Shirai, K., and Honda, M. (**1977**). "Estimation of articulatory motion," in *Dynamic Aspects of Speech Production*, edited by M. Sawashima and F. Cooper (University of Tokyo Press, Tokyo), pp. 279–302.

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

Story, B. H. (**2006**). "A technique for 'tuning' vocal tract area functions based on acoustic sensitivity functions," J. Acoust. Soc. Am. **119**, 715–718.

Story, B. H. (**2005**). "Synergistic modes of vocal tract articulation for American English vowels," J. Acoust. Soc. Am. **118**, 3834–3859.

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

Story, B. H., and Titze, I. R. (**1998**). "Parametrization of vocal tract area functions by empirical orthogonal modes," J. Phonetics **26**, 223–260.

Yehia, H. C., Takeda, K., and Itakura, F. (**1996**). "An acoustically oriented vocal-tract model," IEICE Trans. Inf. Syst. **E79-D**, 1198–1208.

Zheng, Y., Hasegawa-Johnson, M., and Pizza, S. (**2003**). "Analysis of the three-dimensional tongue shape using a three-factor analysis model," J. Acoust. Soc. Am. **113**, 478–486.

Exact wave field simulation for finite-volume scattering problems

Dirk-Jan van Manen

*WesternGeco London Technology Centre, Schlumberger House, Buckingham Gate, Gatwick, West Sussex,
RH6 0NZ, United Kingdom and School of GeoSciences, University of Edinburgh, Grant Institute, West Mains Road,
Edinburgh EH9 3JW, United Kingdom*
dmanen@gatwick.westerngeco.slb.com

Johan O. A. Robertsson

*WesternGeco London Technology Centre, Schlumberger House, Buckingham Gate, Gatwick, West Sussex,
RH6 0NZ, United Kingdom*
jrobertsson@gatwick.westerngeco.slb.com

Andrew Curtis

*ECOSSE (Edinburgh Collaborative of Subsurface Science and Engineering), School of GeoSciences,
University of Edinburgh, Grant Institute, West Mains Road, Edinburgh EH9 3JW, United Kingdom*
andrew.curtis@ed.ac.uk

Abstract: An exact boundary condition is presented for scattering problems involving spatially limited perturbations of arbitrary magnitude to a background model in generally inhomogeneous acoustic media. The boundary condition decouples the wave propagation on a perturbed domain while maintaining all interactions with the background model, thus eliminating the need to regenerate the wave field response on the full model. The method, which is explicit, relies on a Kirchhoff-type integral extrapolation to update the boundary condition at every time step of the simulation. The Green's functions required for extrapolation through the background model are computed efficiently using wave field interferometry.

1. Introduction

Many problems involving wave scattering such as wave form inversion, experimental and industrial design, and nondestructive testing, require evaluation of the wave field response for a suite of closely related models. While these models define material property distributions over some volume V, changes between models may be restricted to a smaller subvolume D_{sct}. Nevertheless, repeated full wave form simulations for each entire model are often a necessity as realistic strong multiple scattering rules out a Born approximation. This situation arises, for example, in complex, multibody scattering problems where only one of the bodies changes shape, material properties, or orientation (Schuster, 1985), or when the changes occur in the vicinity of a free surface or otherwise strong scatterers such as fault planes (Robertsson and Chapman, 2000). We suggest that this paradigm may be broken by combining a Kirchhoff-type integral extrapolation with recent advances in wave field interferometry, resulting in an exact boundary condition for model perturbations of arbitrary magnitude, shape, and size. This allows wave field simulations to be restricted to a subdomain enclosing the changes while retaining all interactions with the full volume.

Nonreflecting boundary conditions based on the Kirchhoff integral were first proposed by Ting and Miksis (1986). By extrapolating the wave field from an artificial surface surrounding a scatterer to the boundary of the computational domain, exact boundary conditions were found such that the computational domain could be truncated without generating spurious boundary reflections. The boundary condition was implemented and tested by Givoli and Co-
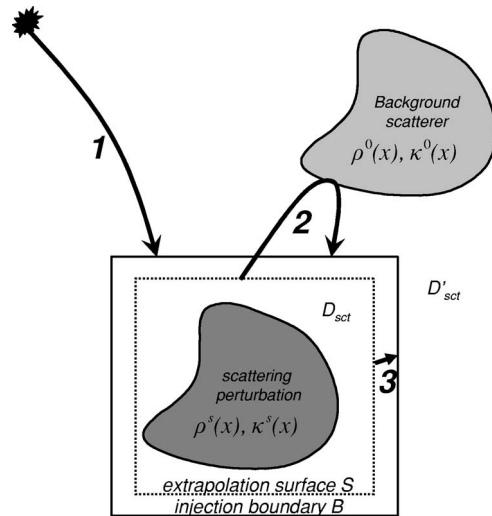
Fig. 1. Definition of the extrapolation surface *S* and injection boundary *B*. Note that three events will contribute to the boundary condition as denoted by arrows: (1) The incident wave field propagating in the background medium, (2) the extrapolated waves that are incoming at the boundary and which radiate back into the subgrid, providing long-range interactions, and (3) the extrapolated waves that are outgoing at the boundary (these will complement the corresponding waves propagating on the subgrid).

hen (1995). Exploring the limiting case where the extrapolation surface coincides with the boundary condition, Teng (2003) obtained a boundary integral equation that could be solved in conjunction with the finite-difference scheme on the subgrid. However, in each case, only exterior (outgoing) wave problems were considered, with nonreflecting boundary conditions outside the scatterer.

On the other hand, if one considers locally perturbed scattering problems, the boundary condition between perturbed and unperturbed domains should treat both incoming and outgoing waves correctly. Schuster (1985) proposes a hybrid boundary integral equation plus Born series modeling scheme that accounts for the long-range interactions between a scattering perturbation and the background by solving a surface boundary integral equation, and which represents the interaction between perturbations using a Born series. However, only homogeneous backgrounds and interiors were considered. Finally, Robertsson and Chapman (2000) describe a method to "inject" a wave field recorded during an initial simulation on the full background model, which then drives the computation on the perturbed interior domain. Their boundary condition correctly accounts for the interaction of the wave field on the subvolume, and between the subvolume and the unperturbed background model, both of which can be arbitrarily heterogeneous. The only part of the wave field missing is that caused by interactions of the altered wave field with the unaltered model outside the subvolume which propagate back into the subvolume and interact with the perturbations again—so-called high-order, long-range interactions (Fig. 1, label 2)

The exact boundary condition presented here combines elements from the "injection" and Kirchhoff extrapolation approaches: It uses an incident wave field as a boundary condition to drive the simulation on the subvolume, but accurately models all high-order, long-range interactions between the perturbed region and the background medium by updating the boundary condition through the evaluation of a Kirchhoff-type integral involving full wave form Green's functions.

One of the main contributions of this letter, and a key enabler for the new method, is the realization that the Green's functions required for extrapolation through the background model can be computed efficiently and flexibly using recent advances in wave field interferom-

etry: By illuminating the model from a surrounding surface with a sequence of conventional forward modeling runs, exact Green's functions between any pair of points can be computed using only cross correlations and summations (Wapenaar, 2004; van Manen *et al.*, 2005).

2. Scattering by an arbitrary inhomogeneous object

It is well known how the wave field scattered by an object with mass density $\rho^s(\mathbf{x})$ and compressibility $\kappa^s(\mathbf{x})$, different from properties $\rho^0(x)$, $\kappa^0(x)$ of the inhomogeneous medium within which it is embedded, originates from the contrast in the material properties. Defining the scattered wave field $\{p^{\text{sct}}, v_k^{\text{sct}}\}$ as the difference between the total wave field $\{p, v_k\}$ propagating in the perturbed model, and the incident wave field $\{p^{\text{inc}}, v_k^{\text{inc}}\}$ propagating in the background model, it is straightforward to show that the scattered wave field quantities satisfy (Fokkema and van den Berg, 1993):

$$\partial_k p^{\text{sct}} + \rho^0 \partial_t v_k^{\text{sct}} = (\rho^0 - \rho^s)\partial_t v_k, \quad \mathbf{x} \in D_{\text{sct}}, \tag{1}$$

$$\partial_k v_k^{\text{sct}} + \kappa^0 \partial_t p^{\text{sct}} = (\kappa^0 - \kappa^s)\partial_t p, \quad \mathbf{x} \in D_{\text{sct}}, \tag{2}$$

where ∂_k and ∂_t denote partial derivatives with respect to the kth spatial dimension ($k \in 1, 2, 3$) and time, respectively, and the Einstein summation convention for repeated indices is used. Equations (1) and (2), show how the scattered wave field originates from body force sources, $(\rho^0 - \rho^s)\partial_t v_k$, and volume injection sources, $(\kappa^0 - \kappa^s)\partial_t p$, acting in the *background* medium. However, the simplicity of Eqs. (1) and (2) is deceptive since the source terms on the right-hand side depend on the unknown *total* wave field quantities p and v_k inside D_{sct}. Nevertheless, if the scattered wave field is known on a surface surrounding the scatterer, ∂D_{sct}, then Eqs. (1) and (2) constitute an acoustic radiation problem and we have the following representation for the scattered pressure at any point, $\mathbf{x}^R$, outside D_{sct} (Fokkema and van den Berg, 1993):

$$p^{\text{sct}}(\mathbf{x}^R, \tau) = \int_0^\tau \int_{\partial D_{\text{sct}}} \left[G^q(\mathbf{x}^R | \mathbf{x}, \tau - t) v_k^{\text{sct}}(\mathbf{x}, t) + \Gamma_k^q(\mathbf{x}^R | \mathbf{x}, \tau - t) p^{\text{sct}}(\mathbf{x}, t) \right] n_k dA dt, \tag{3}$$

where $G^q(\mathbf{x}^R | \mathbf{x}, \tau - t)$ and $\Gamma_k^q(\mathbf{x}^R | \mathbf{x}, \tau - t)$ are the Green's functions for pressure and particle motion due to point sources of volume injection (q in the notation of Fokkema and van den Berg, 1993), respectively, in the *background* medium, ∂D_{sct} is the boundary of D_{sct}, and n_k are the components of the normal to the boundary.

Ting and Miksis (1986) have shown how Eq. (3) can be used to predict outgoing waves arriving at the boundary of a computational domain by extrapolating through free space the scattered wave field from an auxiliary surface surrounding the scatterer to the boundary. This involves substituting free-space Green's functions and evaluating Eq. (3) at time-retarded values $t - r/c$ (with r=distance, c=wave speed, and r/c the travel time between the extrapolation surface and the boundary). Because the waves arriving at the edge of the computational domain are matched by the extrapolated waves, the boundary is nonreflecting.

3. Exact boundary conditions for perturbed scattering problems

If the scatterer occupies just a small part of the background model and the medium is inhomogeneous outside the extrapolation surface, ingoing waves resulting from interaction with the background model will be present and the above-presented approach no longer yields the correct boundary data required to truncate the computational domain. However, Eq. (3) may still be used to extrapolate the wave field to any point outside the extrapolation surface, as long as the exact, full wave form Green's functions for the inhomogeneous background model are used instead of free-space Green's functions. Multiple scattering between the perturbed region and the inhomogeneous background model of the radiated wave field may then affect the boundary data at all later times. We write Eq. (3) recursively to make the contribution of the scattered wave field at time t to all later times explicit. After discretizing the convolution integral in time this gives

$$\hat{p}^{\text{sct}}(\mathbf{x}^R, l, n) = \hat{p}^{\text{sct}}(\mathbf{x}^R, l, n-1) + \int_{\partial D_{\text{sct}}} [\hat{G}^q(\mathbf{x}^R|\mathbf{x}, l-n)\hat{v}_k(\mathbf{x}, n) + \hat{\Gamma}_k^q(\mathbf{x}^R|\mathbf{x}, l-n)\hat{p}(\mathbf{x}, n)]n_k dA,$$

$$(4)$$

where a caret is used to differentiate between continuous time and sampled quantities. Note that the integral over t in Eq. (3) is implicit in the recursion in Eq. (4) and that the discrete-time indices l and n correspond to τ and t, respectively. Thus, to update the scattered wave field $\hat{p}^{\text{sct}}(\mathbf{x}^R, l, n-1)$ at $\mathbf{x}^R$ at time step n of the computation for all future time steps $l > n$, one has to scale the Green's functions $\hat{G}^q(\mathbf{x}^R|\mathbf{x}, l-n)$ and $n_k\hat{\Gamma}_k^q(\mathbf{x}^R|\mathbf{x}, l-n)$ by the current value of the normal component of particle velocity $n_k, \hat{v}_k(\mathbf{x}n)$, and the pressure $\hat{p}(\mathbf{x}, n)$ on the extrapolation surface, respectively, and add this to the previously computed values at $\mathbf{x}^R$. Equation (4) needs to be complemented by the incident wave field $\hat{p}^{\text{inc}}(\mathbf{x}^R, n)$ to give the total wave field at $\mathbf{x}^R$. The resulting boundary condition is exact and equivalent to the Neumann series solution to the scattering problem (Snieder, 2002): It includes all orders of interactions between the background model and the perturbations.

To implement a wave field simulation on a limited region, Eq. (4) is used to update the boundary condition at each point on the boundary of the truncated domain, using the scattered wave field emitted from the perturbed domain at each time step. As in the method of Ting and Miksis, outgoing waves are absorbed at the boundary, because they are matched by the boundary condition. However, Eq. (4) also generates the desired incoming waves from higher order interactions with the background model, which are subsequently radiated inwards into the subgrid as desired. We now show how to compute the Green's functions in Eq. (4) required for extrapolation efficiently.

4. Interferometry

In the interferometric method, waves at two receiver locations are correlated to find the Green's function between them (Weaver and Lobkis, 2001). Recently, it was shown that there is a strong link between interferometry and reciprocity. Consider the acoustic reciprocity theorem of the correlation type (Fokkema and van den Berg, 1993);

$$\int_{\mathbf{x} \in \partial D} [C_t\{p^A, v_k^B\} + C_t\{v_k^A, p^B\}]n_k dA = \int_{\mathbf{x} \in D} [C_t\{f_k^A, v_k^B\} + C_t\{p^A, q^B\} + C_t\{v_k^A, f_k^B,\}$$

$$+ C_t\{q^A, p^B\}]dV, \qquad (5)$$

where $C_t\{\cdot\}$ denotes temporal cross correlation. An interferometric representation for the pressure due to a point source of volume injection, $G^q(\mathbf{x}^A|\mathbf{x}^B, t)$, between points $\mathbf{x}^A$ and $\mathbf{x}^B$, can be derived by taking as state A the wave field generated by a point source of volume injection at $\mathbf{x}^A$: $\{p^A, v_k^A\}(\mathbf{x}, t) = \{G^q, \Gamma_k^q\}(\mathbf{x}|\mathbf{x}^A, t)$ and $\{q^A, f_k^A\}(\mathbf{x}, t) = \{\delta(t)\delta(\mathbf{x} - \mathbf{x}^A), 0\}$ and state B the wave field generated by a point source of volume injection at x^B: $\{p^B, v_k^B\}(\mathbf{x}, t) = \{G^q, \Gamma_k^q\}(\mathbf{x}|\mathbf{x}^B, t)$ and $\{q^B, f_k^B\}(\mathbf{x}, t) = \{\delta(t)\delta(\mathbf{x} - \mathbf{x}^B), 0\}$.

Inserting these expressions into Eq. (5), performing the volume integrations, and using reciprocity $[G^q(\mathbf{x}^A|\mathbf{x}^B, t) = G^q(\mathbf{x}^B|\mathbf{x}^A, t)$ and $\Gamma_k^q(\mathbf{x}|\mathbf{x}^B, t) = -G_k^f(\mathbf{x}^B|x, t)]$ we find:

$$G^q(\mathbf{x}^B|\mathbf{x}^A,t) + G^q(\mathbf{x}^B|\mathbf{x}^A,-t) = -\int_{\mathbf{x}\in\partial D}[G^q(\mathbf{x}^A|\mathbf{x},t) * G_k^f(\mathbf{x}^B|\mathbf{x},-t)$$

$$+ G_k^f(\mathbf{x}^A|\mathbf{x},t) * G^q(\mathbf{x}^B|\mathbf{x},-t)]n_k dA, \qquad (6)$$

where as asterisk denotes temporal convolution and $G_k^f(\mathbf{x}^B|x,t)$ denotes the pressure in $\mathbf{x}^B$ due to a unidirectional point source force in the k direction at $\mathbf{x}$. Similarly, an interferometric representation for the pressure due to a point force source can be derived. By systematically illuminating a model from the surrounding surface, while storing the wave field in as many points in the interior as possible, full wave form Green's functions can be computed for any pair of points using only cross correlation and numerical integration (van Manen *et al.*, 2005; 2006). This allows the Green's functions required to update the boundary condition to be computed efficiently and flexibly for any subdomain, once the initial illumination has been stored everywhere.

5. One-dimensional example

The exact boundary condition is demonstrated in an example using a staggered finite-difference approximation of the one-dimensional acoustic wave equation. The model consists of a single scattering layer (propagation velocity $c_s = 1750$ m/s, mass density $\rho_s = 1250$ kg/m^3) embedded in a homogeneous background medium between 130 m and 170 m depth ($c_0 = 2000$ m/s, $\rho_0 = 1000$ kg/m^3) and with a free surface at the top.

Since the model is one dimensional and bounded by a free surface at the top, a single source at the bottom of the well is sufficient to illuminate the model completely. Thus, only two conventional forward modeling runs were performed (one for each source type) and the data stored at every gridpoint. Nonreflecting boundary conditions were used just below the source to truncate the computational domain. Given the data of these two initial simulations, Green's functions between arbitrary points in the well can be computed using Eq. (6).

An incident wave field was calculated using interferometry, for a volume injection source at 50 m depth, and receivers collocated with the pressure points at the edge of the planned truncated computational domain. Since the finite-difference calculations are done on a staggered grid, whereas the Kirchhoff integral (and also the integral in the interferometric construction) is evaluated for the pressure and particle velocity quantities collocated in space and time, care should be taken that the required pressure and particle velocities are linearly interpolated to the same location and time.

Auxiliary extrapolation "surfaces" were defined just above and below the perturbation at 125 and 175 m depth, corresponding to boundary S in Fig. 1. Next, the model was significantly perturbed by increasing the velocity by 500 m/s and the density by 250 kg/m^3 (both changes greater than 25%) in the scatterer. Since there is a free surface, waves scattering off the perturbation will reflect at the free surface and repeatedly interact with the perturbation. Thus, high-order, long-range interactions will be present, ruling out a Born approximation or conventional finite-difference injection to compute the response on the perturbed model.

To compute the response using the new methodology, the computational domain was truncated 15 m above and below the extrapolation points (at 110 and 190 m depth, respectively) corresponding to boundary B in Fig. 1. The offset of 15 m between the extrapolation surface and the boundary of the truncated domain was chosen to prevent errors due to the diffraction limit inherent in the interferometric Green's functions (de Rosny and Fink, 2002). Without loss of generality, we opted for a pressure (Dirichlet) boundary condition at the edge of the truncated computational domain and collocated the evaluation points of the Kirchhoff integral with the (staggered) *FD* pressure points at the edge of the grid. Thus, Green's functions for extrapolation through the background model need to be computed between all combinations of points with one point on the extrapolation and one point on the "injection" surface (i.e., between 125 and 110 m, 125 and 190 m, 175 and 110 m, and 175 and 190 m) and for both pressure-to-pressure and particle velocity-to-pressure interactions, giving a total of eight extrapolation Green's functions for this simple one-dimensional example.
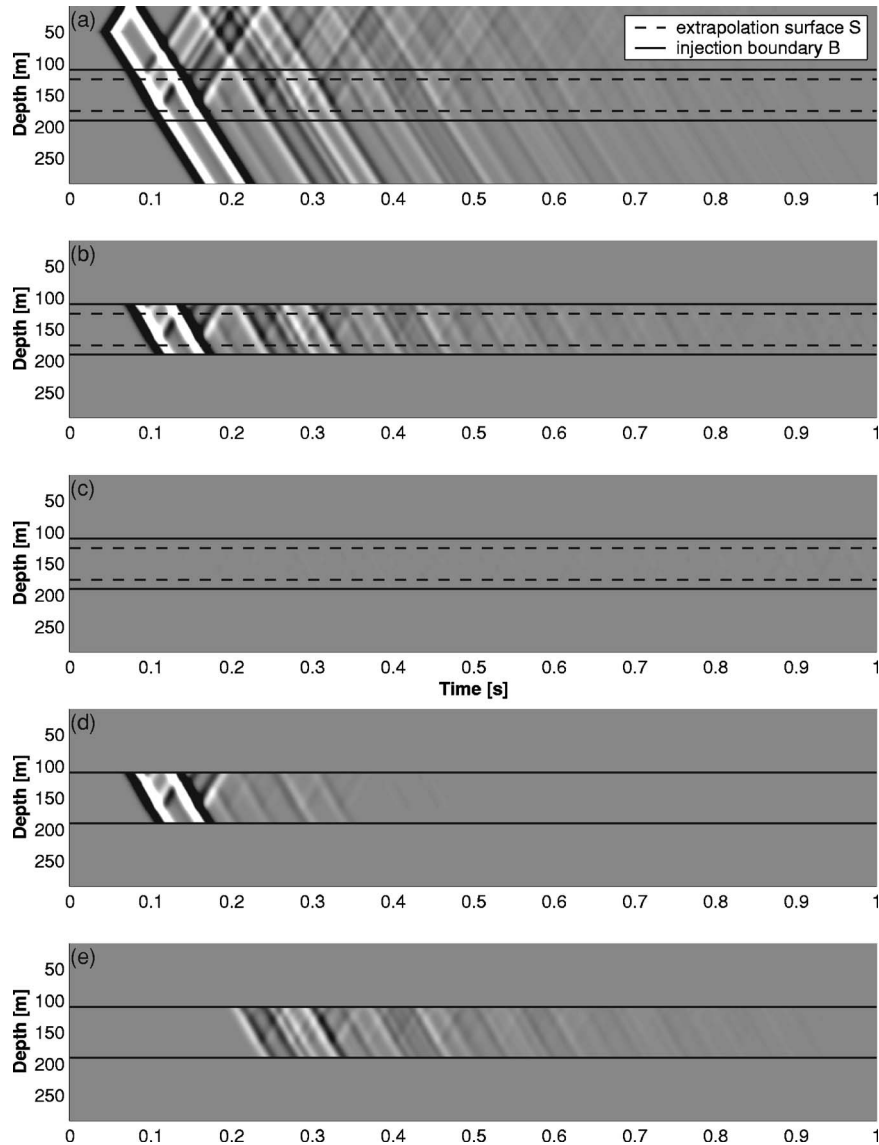
Fig. 2. Comparison of the proposed exact boundary condition and conventional *FD* injection to a directly computed reference. (a) Pressure, directly computed for the perturbed model by *FD* on the full grid. (b) Pressure, computed using the new method on the subgrid *D*. (c) Difference between (a) and (b) for the extent of the subgrid (110–190 m). (d) Pressure on the subgrid D, computed using conventional *FD* injection. (e) Difference between (a) and (d) for the extent of the subgrid (110–190 m).

In practice, the Kirchhoff extrapolation is evaluated for every time step of the finite-difference simulation on the truncated perturbed domain and the resulting seismograms used to update the buffer of future boundary values [Eq. (4)]. The next sample from the buffer is then used as the *BC* for the subsequent time step in the *FD* calculation. The wave field on the perturbed grid resulting from the new boundary condition is shown in Fig. 2(b). Note that no additional absorbing boundary conditions were used outside the new exact boundary condition. The resulting pressure wave field can be compared to a reference wave field in Fig. 2(a) calculated by a full *FD* simulation across the entire domain. In Fig. 2(c), the difference between Figs.

2(a) and 2(b) is shown. The first high-order long-range interactions start between 0.2 and 0.24 s (depending on the depth). Thus, the high-order long-range interactions are reproduced exactly.

The main advance embodied in this method is to include all high-order, long-range interactions between the scattered wave field and the background medium. We therefore compare the results in Figs. 2(b) and 2(c) with corresponding results using the wave field injection method of Robertsson and Chapman (2002), shown in Figs. 2(d) and 2(e). That method only includes first-order, long-range interactions and hence, while the wave field simulated on the subdomain matches well at times up to 0.2 s, the arrival of the first, higher-order interaction thereafter causes errors in the local simulation. These are the errors that are removed by the new method.

6. Conclusion

The presented boundary condition is exact and can be used to compute the response, including all high-order, long-range interactions, to arbitrary perturbations in inhomogeneous models. There are no restrictions on the medium between the extrapolation surface and the boundary of the truncated computational domain as long as it is exactly the same as in the background model. No additional absorbing boundaries are necessary and no special functions need to be evaluated. It is possible to extend the method to cases where multiple, distinct subdomains of the model have been perturbed. In this case, the main result [Eq. (4)] is still valid but it should be used to update the boundary conditions at all boundary points around all subdomains at each time step, with wave fields extrapolated from boundary ∂D_{sct}, which then spans the extrapolation surfaces of all subdomains. Thus, the presented approach also explicitly models the cross interactions between different finite-volume scattering regions. A key enabling feature of the new method is that the Green's functions required for extrapolation through the background model can be computed efficiently and flexibly using wave field interferometry. Finally, in principle the method can be extended to electromagnetic and elastic wave propagation.

de Rosny, J., and Fink, M. (**2002**). "Overcoming the diffraction limit in wave physics using a time-reversal mirror and a novel acoustic sink," Phys. Rev. Lett. **89**, 124301(1)-(4).

Fokkema, J. T., and van den Berg, P. M. (**1993**). *Seismic Applications of Acoustic Reciprocity* (Elsevier, New York).

Givoli, D., and Cohen, D. (**1995**). "Nonreflecting boundary conditions based on Kirchhoff-type formulae," J. Comput. Phys. **117**, 102–113.

Robertsson, J. O. A., and Chapman, C. H. (**2000**). "An efficient method for calculating finite-difference seismograms after model alterations," Geophysics **65**, 907–918.

Schuster, G. T. (**1985**). "A hybrid BIE+Born series modeling scheme: Generalized Born series," J. Acoust. Soc. Am. **77**, 865–879.

Snieder, R. (**2002**). "General theory of elastic wave scattering," in *Scattering and Inverse Scattering in Pure and Applied Science* (Academic, New York), pp. 528–542.

Teng, Z. (**2003**). "Exact boundary condition for time-dependent wave equation based on boundary integral," J. Comput. Phys. **190**, 398–418.

Ting, L., and Miksis, M. J. (**1986**). "Exact boundary conditions for scattering problems," J. Acoust. Soc. Am. **80**, 1825–1827.

van Manen, D., Robertsson, J. O. A., and Curtis, A. (**2005**). "Modeling of wave propagation in inhomogeneous media," Phys. Rev. Lett. **94**, 164301(1)-(4).

van Manen, D. J., Curtis, A., and Robertsson, J. O. A. (**2006**). "Interferometric modeling of wave propagation in inhomogeneous elastic media using time reversal and reciprocity," Geophysics **71**, SI47–SI60.

Wapenaar, K. (**2004**). "Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation," Phys. Rev. Lett. **93**, 254301(1)-(4).

Weaver, R. L., and Lobkis, O. I. (**2001**). "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301(1)-(4).

Nonperiodicity of the flow within the gap of a thermoacoustic couple at high amplitudes

Arganthaël Berson and Philippe Blanc-Benon
LMFA, UMR CNRS 5509, Ecole Centrale de Lyon, 69134 Ecully Cedex, France
arganthael.berson@ec-lyon.fr, philippe.blanc-benon@ec-lyon.fr

Abstract: The flow inside a thermoacoustic couple is investigated experimentally using particle image velocimetry. Measurements show the oscillation of the shear layers flowing out of a single stack, thus forming an asymmetric vortex street at high driving amplitudes. Development of vortices is also observed within the gap of a thermoacoustic couple. It causes the flow not to repeat from one acoustic period to another. The nonperiodicity of the flow will lead to unsteady heat transfer between the stack and heat exchangers and to the oscillation of the cooling load.

1. Introduction

The design of appropriate heat exchangers is a crucial issue for the development of efficient thermoacoustic systems. Recent numerical studies (Marx, 2003; Besnoin and Knio, 2004) have shown that heat transfer through the gap between a thermoacoustic stack and a heat exchanger is strongly coupled with the flow patterns in this area. The improvement of thermoacoustic system performances requires a better understanding of such flows, especially at high amplitudes when nonlinear effects become more important (Poese and Garrett, 2000). The flow behind a single stack has been investigated experimentally by Wetzel and Herman (2000), Duffourd (2001), and Berson *et al.* (2007), showing vortex generation close to the end of the stack plates at low amplitude. Such flow is responsible for minor losses and energy dissipation as was explored by Waxler (2001), Wakeland and Keolian (2002), and Smith and Swift (2003).

After a brief description of the experimental setup, this study presents particle image velocimetry (PIV) measurements behind a single stack. Vortex shedding at high acoustic pressure level is reported and dimensionless numbers are estimated. The last part investigates the flow between two adjacent stacks separated by a gap as a model for a stack coupled to a heat exchanger. The results show that, at high amplitudes, vortex shedding causes the flow in the gap not to be repeatable from one acoustic period to another. A direct consequence will be the oscillation of the cooling load, as was demonstrated in the numerical studies by Besnoin (2001).

2. Experimental setup

2.1 The thermoacoustic refrigerator

Experiments are conducted in a quarter-wavelength standing-wave thermoacoustic refrigerator. The resonator is a closed cylindrical tube of length $L=150$ mm and diameter $2R=30$ mm. A 25-mm-long part of the resonator is made of specially polished glass in order to ensure high quality optical access for PIV measurements. The system is driven by a GELEC EDM8760F loudspeaker that is connected to the resonator by an exponential adaptation horn. The fluid inside the resonator is air at atmospheric pressure. The driver delivers acoustic pressure levels up to $P_{ac}=5000$ Pa at the acoustic resonance frequency $f_{res} \approx 460$ Hz. Acoustic pressure is monitored by a 1/4 in. Bruel & Kjaer microphone flush mounted at the closed end of the resonator.

Thermoacoustic stacks are hand-made of thin glass plates separated by plastic spacers. Three different types of stacks have been built, which are hereafter referred to as stacks A, B,

Table 1. Characteristics of the different stack configurations. l is the plate length, y_0 is the plate spacing, e_0 is the plate thickness, BR $= y_0/(y_0+e_0+S_{solid})$ is the blockage ratio, S_{solid} is the area of the support and $\delta_\nu = \sqrt{\nu/(\pi f_{res})}$ is the viscous boundary layer thickness, with ν the kinematic viscosity of the fluid.

Stack	A	B	C	D	E
l (mm)	18	18	24	25	$\cdots$
y_0 (mm)	0.41	0.27	0.42	1	0.93
e_0 (mm)	0.17	0.17	0.17	1	0.47
BR	0.39	0.34	0.62	0.5	0.66
f_{res} (Hz)	461	455	464	210	110
δ_ν (mm)	0.10	0.10	0.10	0.15	0.21

and C. Their geometrical characteristics are summarized in Table 1. Data for stacks D and E, which are also reported in Table 1, are taken from the experimental work presented in Berson et al. (2007) and from the numerical simulations of Besnoin and Knio (2004). Further details about configurations D and E can be found in these references. In the present configurations, the hot end of the stack is located at $x_h = 65$ mm (x_h being the distance from the closed end of the resonator). Additional tests were performed with stack A located at $x_h = 115$ mm. The investigation of the flow between two coupled stacks refers to two type-A stacks. In this case, a first stack is located at $x_h = 65$ mm and a second one is placed just behind and closer to the closed end of the resonator. Care was taken to make the plates of both stacks as parallel as possible with respect to each other. Adequate plastic spacers separate both stacks, ensuring gap width of approximately $g = 0.7$, 1.3, or 2.3 mm between the plate ends. Although these locations are not optimal for the thermoacoustic effect, we believe that the general behavior of the flow hardly depends on the temperature gradient along the stack and is mostly aerodynamical.

2.2 The PIV measuring system

Measurements are performed with a LaVision™ digital camera and a $12\times$ Navitar zoom lens. The measurement area is in a plane perpendicular to the stack plates and parallel to the resonator axis. Its size ranges from 10×8 mm^2 down to 2.5×3 mm^2. A dual-resonator Nd:YAG laser illuminates the flow field that is seeded with droplets of parrafin oil smoke. Data acquisition is synchronized with the signal driving the loudspeaker and we perform phase averaging to overcome the sampling frequency limitations of the acquisition system. An acoustic period is decomposed in 32 equally spaced phases Φ, where $\Phi = 0°$ corresponds to the instant when acoustic pressure is zero and velocity is maximal, the flow being directed toward the loudspeaker. For each selected phase, velocity fields are averaged over 50 instant measurements. A more detailed description and the validation of this PIV measurement technique can be found in Duffourd (2001), Blanc-Benon et al. (2003), and Berson et al. (2007). For analysis purposes, we calculate two-dimensional vorticity fields Ω and $\Gamma_2 = 2/\pi$ contours from phase-averaged velocity measurements. Γ_2 function is a nondimensional quantity defined at a point M of the space as

$$\Gamma_2(\mathbf{x}) = \frac{1}{S} \int \int_{M' \in S} \sin\theta \mathrm{d}x_i' \mathrm{d}x_j', \qquad (1)$$

with S the integration domain, centered on M. M' belongs to S and its coordinates are $\mathbf{x}'$. θ is the angle between $\mathbf{x}' - \mathbf{x}$ and $\mathbf{u}(M') - \mathbf{u}(M)$, $\mathbf{u}(M)$ being the velocity at coordinate M. Subscripts i and j refer to the vector components in the plane. Γ_2 function is comprised between $2/\pi$ and 1 when the flow is locally dominated by rotation, thus allowing one to detect vortex boundaries. A complete description of Γ_2 function is given in Graftieaux et al. (2001).
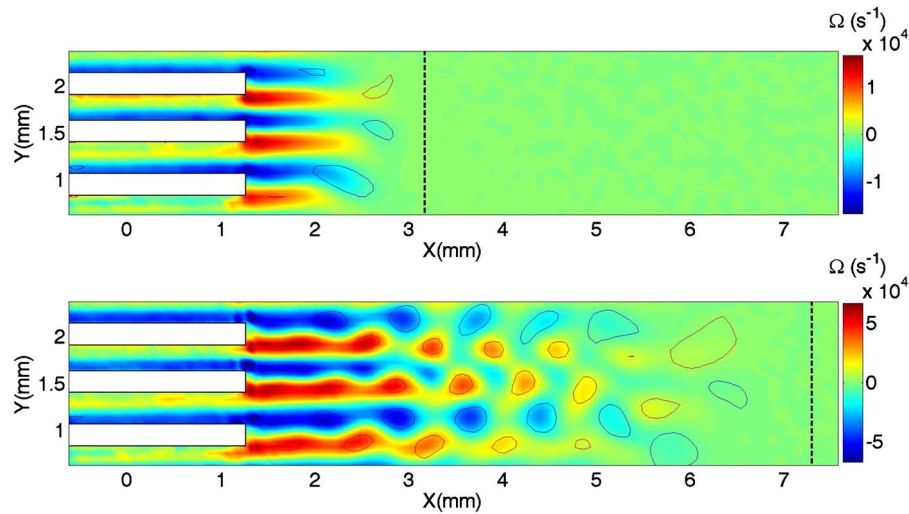
Fig. 1. (Color online) Vorticity fields $\Omega(s^{-1})$ and $\Gamma_2 = 2/\pi$ contours behind stack A. $\Phi = 225°$. Top panel: P_{ac} = 1000 Pa, bottom panel: P_{ac} = 4000 Pa. The black dashed line represents the distance $2d_{ac}$ away from the plate edge. Note that the color scales are different for improved readability.

3. Results and discussion

3.1 Oscillation of the flow behind a single stack at high amplitudes

The flow field behind a single stack has been previously investigated experimentally by Duffourd (2001) and Berson et al. (2007) for low acoustic pressure level configurations ($P_{ac} \leqslant 2000$ Pa). Vortex shedding was observed in Berson et al. (2007) in the case of thick stack plates ($e_0/\delta_\nu \approx 7$). With more practical and thinner plates ($e_0/\delta_\nu < 2$), as is the case here, no vortex shedding was observed in the range of acoustic velocity investigated at the time.

Figure 1 shows vorticity fields and $\Gamma_2 = 2/\pi$ contours of the flow behind stack A at phase $\Phi = 225°$ for $P_{ac} = 1000$ Pa and $P_{ac} = 4000$ Pa. Measurements are phase-averaged over 50 images. A movie of the complete acoustic period is available in Mm. 1. Note that due to a slight shift in the laser alignment with the stack plates, the flow within the channels of the stack is not measured accurately. Such flow is detailed in Duffourd (2001) and Berson et al. (2007). The measurements of the flow behind the stack are not affected by this shift.

> Mm1: Vorticity fields Ω (s^{-1}) and $\Gamma_2 = 2/\pi$ contours behind stack A along an acoustic period (from $\Phi = 0°$ to 360°). Sampling frequency: $1/22.5°$ except between $\Phi = 135°$ and 315° where sampling frequency is doubled. Top: $P_{ac} = 1000$ Pa, bottom: $P_{ac} = 4000$ Pa. Note that the color scales are different for improved readability. The black dashed line represents the distance $2d_{ac}$ away from the plate edge. This is a file of type "avi" (5.9 Mbytes).

Acoustic velocity amplitudes $u_{ac}(x_h)$ corresponding to $P_{ac} = 1000$ Pa and $P_{ac} = 4000$ Pa at the stack location $x_h = 115$ mm are $u_{ac}(x_h) = 2$ m s^{-1} and $u_{ac}(x_h) = 8$ m s^{-1}, respectively. At $P_{ac} = 1000$ Pa, vorticity layers extend out of the stack during the expulsion phases (when the fluid is flowing out of the hot side of the stack), thus prolongating the oscillating boundary layers that develop along the stack plates. At $P_{ac} = 4000$ Pa, we observe that the vorticity layers outside the stack oscillate and generate asymmetric streets of counter-rotating vortex pairs shedding away from the stack. In the range of acoustic velocities studied here, oscillation of the vorticity layers occurs only during the phases when the fluid flows out of the stack and decelerates due to the acoustic standing wave (from $\Phi = 180°$ to $\Phi = 270°$). Note that gravity is directed perpendicular to the measurement plane so that buoyancy should not be responsible for these oscillations. We also notice that, in all the configurations tested in this study, the vor-

Table 2. Reynolds number and formation number at onset of vortex shedding and Strouhal numbers based on various characteristic lengths for different stack configurations.

Stack	A $(x_h=65$ mm$)$	A $(x_h=115$ mm$)$	B	C	D	E
$\mathrm{Re_{onset}}$	94	99	108	60	102	$\cdots$
T^{*}_{onset}	6.4	6.8	9.4	6.0	11.2	$\cdots$
St_{δ_ν}	0.12	0.12	0.14	0.11	0.07	0.10
St_{e_0}	0.20	0.20	0.25	0.19	0.44	0.23
St_{y_0}	0.48	0.48	0.39	0.46	0.44	0.46

ticity layers never extend further than two acoustic displacements away from the stack edge (indicated by a dashed line in Fig. 1). An acoustic displacement $d_{\mathrm{ac}}(x)=u_{\mathrm{ac}}(x)/\omega$ is the distance a fluid particle can travel during an acoustic half-period.

Dimensionless numbers are calculated in order to provide a better insight into the phenomenon of vortex shedding behind the stack. We define a Reynolds number Re based on viscous boundary layer thickness and velocity amplitude inside the channels of the stack, as proposed by Merkli and Thomann (1975):

$$\mathrm{Re}=\sqrt{2}\frac{u_{\mathrm{ac}}}{\mathrm{BR}}\frac{\delta_\nu}{\nu}, \tag{2}$$

where ν is the kinematic viscosity of air. The viscous boundary layer thickness δ_ν and blockage ratio BR are given in Table 1 for each configuration. Merkli and Thomann (1975) found that onset of turbulence in oscillating flows occurs when $\mathrm{Re}\geq400$ approximately. The highest Reynolds number achieved during these experiments is $\mathrm{Re}\approx198$. As it is well below the critical Reynolds number the flow remains laminar, even though vortex shedding occurs. Reynolds numbers corresponding to the onset of vortex shedding $\mathrm{Re_{onset}}$ are presented in Table 2. $\mathrm{Re_{onset}}$ is the Reynolds number calculated at the lowest acoustic pressure level that was tested for which vortex shedding occurs. We consider that shedding occurs when at least two consecutive pairs of counter-rotating vortices detach from the stack. According to the collected data, there is no critical Reynolds number value for the onset of oscillation.

Different Strouhal numbers are also presented in Table 2. Strouhal numbers are defined as

$$\mathrm{St}_D=\frac{f_{\mathrm{shed}}D}{u_{\mathrm{conv}}}=\frac{D}{\Delta x}, \tag{3}$$

where f_{shed} is the shedding frequency, u_{conv} is the convection velocity of vortices, Δx is the distance between two consecutive pairs of vortices, and D is one of the characteristic lengths δ_ν, e_0 and y_0. For each configuration, Strouhal numbers are averaged over all the phase-averaged velocity fields corresponding to this configuration. We observe that the Strouhal number does not depend on the Reynolds number for a given stack configuration. It is noticeable that Strouhal numbers calculated with $D=y_0$ take very close values, $\mathrm{St}_{y_0}\approx0.44\pm10\%$, for the different configurations tested in this study.

Another approach would be to estimate the formation number $T^{*}_{\mathrm{onset}}=2d_{\mathrm{ac}}/y_0$ corresponding to the onset of vortex shedding. Indeed, in the case of vortex rings, it has been demonstrated that the detachment of the head vortex from the outflowing jet always occurs at $T^{*}=4$ (Gharib et al., 1998; Zhao et al., 2000). But, from the measurements reported in Table 2, no such universal value could be found. A more extensive study covering a larger range of geometries would be required to obtain more general laws between the geometry of the stack and the flow behind it, and possibly define a criterion for the onset of shedding.

3.2 Loss of periodicity of the flow in the gap between two facing stacks due to vortex shedding

An analytical model (Gusev et al., 2000) and recent computational studies (Marx and Blanc-Benon, 2004; Besnoin and Knio, 2004) have demonstrated that a gap between the stack and
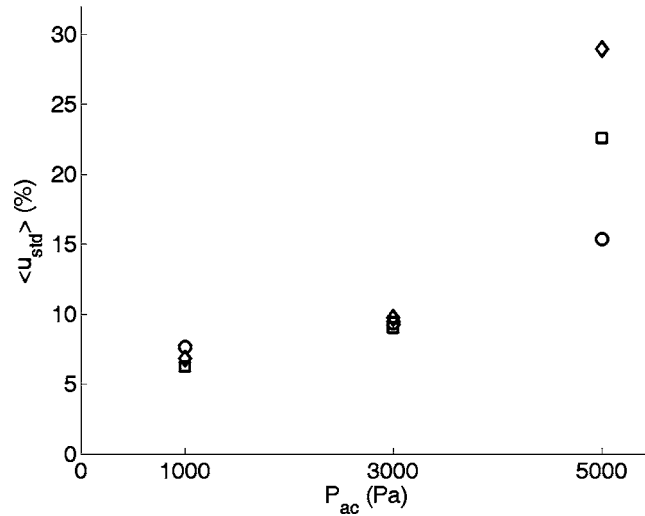
Fig. 2. Standard deviation averaged along an acoustic period and over the gap area at different acoustic pressure levels. ○: $g=0.7$ mm, □: $g=1.3$ mm, and ◇: $g=2.3$ mm.

both heat exchangers (hereafter referred to as "the gap") is favorable to the performances of the thermoacoustic system. Mm. 2 shows vorticity fields and $\Gamma_2=2/\pi$ contours of the flow in the gap along an acoustic period, for four cases corresponding to $P_{ac}=1000$ Pa with gap widths $g/(2d_{ac})=0.8$ and 2.6, and $P_{ac}=5000$ Pa with gap widths $g/(2d_{ac})=0.2$ and 0.5. The hot side of the stack represented on the left in Mm. 2 is located at $x_h=65$ mm. Measurements are phase-averaged over 50 images.

> Mm2: Vorticity fields Ω (s^{-1}) and $\Gamma_2=2/\pi$ contours in the gap along an acoustic period (from Φ =0° to 360°). Sampling frequency: 1/22.5° except between $\Phi=135°$ and 315° where sampling frequency is doubled. Top: $P_{ac}=1000$ Pa, bottom: $P_{ac}=5000$ Pa, left: $g=0.7$ mm, right: g =2.3 mm. Note that the color scales are different for improved readability. The black dashed line represents the distance $2d_{ac}$ away from the plate edge. This is a file of type "avi" (5.9 Mbytes).

At $P_{ac}=1000$ Pa, vortex shedding does not occur. When $g>2d_{ac}$, a fluid particle flowing outside one stack never reaches the other stack and the flow is not disturbed by the presence of another stack. When $g<2d_{ac}$, the boundary layers flowing from one stack channel extend through the gap toward the facing channel of the other stack. The flow behaves as if there were virtual stack plates in the gap. At $P_{ac}=5000$ Pa, vortex shedding occurs in the gap and the flow is similar to the flow behind a single stack described in Sec. 3.1. As Besnoin and Knio (2004) have shown, the disturbance of the flow within the gap will affect heat transfer between the stack and the heat exchangers.

We calculate the relative standard deviation of velocity measurements u_{std} that we define as

$$u_{std}(x) = \frac{1}{u_{ac}(x)}\left[\frac{1}{N}\sum_{n=1}^{N}(\bar{u}(x)-u_n(x))^2\right]^{1/2}, \tag{4}$$

where $\bar{u}(x)$ is the phase-averaged velocity calculated over $N=50$ instant velocity measurements $u_n(x)$ performed for each phase. By definition, since measurements are phase-locked, a perfectly periodic phenomenon has zero standard deviation. Figure 2 shows velocity relative standard deviation averaged along an acoustic period and over the gap area $\langle u_{std}(x_h)\rangle$ for three different acoustic pressure levels and various gap widths. Relative standard deviation is approximately 6.5% and 9.5% at $P_{ac}=1000$ Pa and $P_{ac}=3000$ Pa, respectively. The flow is quite repeatable from one acoustic period to another at low acoustic pressure level, regardless of the

gap width. At high amplitudes $P_{ac}=5000$ Pa, relative standard deviation increases with gap width, attaining $\langle u_{std}(x_h)\rangle=29\%$ for $g/2d_{ac}=0.5$. The flow is no longer periodic. This confirms the results obtained numerically by Besnoin (2001). As the flow within the gap does not repeat from one acoustic period to another, the average heat transfer through the gap is no longer steady, and oscillation of the cooling load will occur at high driving amplitudes.

4. Conclusion

This study shows the nonperiodic character of the flow within the gap between two coupled stacks as a model for a stack-heat exchanger couple in a thermoacoustic refrigerator driven at high amplitudes. PIV measurements are first conducted behind a single stack where vortex shedding is observed at high acoustic pressure level. The estimation of dimensionless numbers highlights instructive features of this phenomenon that should help in defining a criterion for the onset of vortex shedding in further studies. Standard deviation of velocity measurements shows that the flow between two coupled stacks is not repeated from one acoustic period to another at high acoustic pressure level. This loss of periodicity will lead to the oscillation of cooling load, as Besnoin (2001) has shown. Future works will include a more extensive study of the onset of oscillation of vorticity layers behind the stack and heat flux measurements on heat exchangers.

References and links

Berson, A., Michard, M., and Blanc-Benon, Ph. (**2007**). "Measurement of acoustic velocity in the stack of a thermoacoustic refrigerator using Particle Image Velocimetry," Heat and Mass Transfer, in press.

Besnoin, E. (**2001**). "Numerical study of thermoacoustic heat exchangers," Ph.D. thesis, The Johns Hopkins University, Baltimore, MD.

Besnoin, E., and Knio, O. (**2004**). "Numerical study of thermoacoustic heat exchangers," Acta. Acust. Acust. **90**, 432–444.

Blanc-Benon, Ph., Besnoin, E., and Knio, O. (**2003**). "Experimental and computational visualization of the flow field in a thermoacoustic stack," C. R. Mec. **331**, 17–24.

Duffourd, S. (**2001**). "Réfrigérateur thermoacoustique: Études analytiques et expérimentales en vue d'une miniaturisation," Ph.D. thesis. 2001-06, Ecole Centrale de Lyon (in French), Ecully, France. ("Thermoacoustic refrigerator: Analytical and experimental studies with the prospect of miniature systems".)

Gharib, M., Rambod, E., and Shariff, K. (**1998**). "A universal time scale for vortex ring formation," J. Fluid Mech. **360**, 121–140.

Graftieaux, L., Michard, M., and Grosjean, N. (**2001**). "Combining PIV, POD and vortex identification algorithms for the study of unsteady turbulent swirling flows," Meas. Sci. Technol. **12**, 1422–1429.

Gusev, V., Lotton, P., Baillet, H., Job, S., and Bruneau, M. (**2000**). "Relaxation-time approximation for analytical evaluation of temperature field in a thermoacoustic stack," J. Sound Vib. **235**, 711–726.

Marx, D. (**2003**). "Simulation numérique d'un réfrigérateur thermoacoustique," Ph.D. thesis. 2003-34, Ecole Centrale de Lyon (in French), Ecully, France. ("Numerical simulation of a thermoacoustic refrigerator".)

Marx, D., and Blanc-Benon, Ph. (**2004**). "Numerical simulation of the stack-heat exchangers coupling in a thermoacoustic refrigerator," AIAA J. **42**, 1338–1347.

Merkli, P., and Thomann, H. (**1975**). "Transition to turbulence in oscillating pipe flow," J. Fluid Mech. **68**, 567–575.

Poese, M., and Garrett, S. (**2000**). "Performance measurements on a thermoacoustic refrigerator driven at high amplitudes," J. Acoust. Soc. Am. **107**, 2480–2486.

Smith, B., and Swift, G. (**2003**). "Power dissipation and time-averaged pressure in oscillating flow through a sudden area change," J. Acoust. Soc. Am. **113**, 2455–2463.

Wakeland, R., and Keolian, R. (**2002**). "Influence of velocity profile nonuniformity on minor losses for flow exiting thermoacoustic heat exchangers," J. Acoust. Soc. Am. **112**, 1249–1252.

Waxler, R. (**2001**). "Stationary velocity and pressure gradients in a thermoacoustic stack," J. Acoust. Soc. Am. **109**, 2739–2750.

Wetzel, M., and Herman, C. (**2000**). "Experimental study of thermoacoustic effect on a single plate. I. Temperature fields," Heat Mass Transfer **36**, 7–20.

Zhao, W., Frankel, S., and Mongeau, L. (**2000**). "Effects of trailing edge instability on vortex ring formation," Phys. Fluids **12**, 589–596.

Use of a sigmoidal-shaped function for noise attenuation in cochlear implants

Yi Hu, Philipos C. Loizou,[a] Ning Li, and Kalyan Kasturi
Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas 75083-0688
yihuyxy@utdallas.edu, loizou@utdallas.edu, lining@student.utdallas.edu, kalyan@utdallas.edu

Abstract: A new noise reduction algorithm is proposed for cochlear implants that applies attenuation to the noisy envelopes inversely proportional to the estimated signal-to-noise ratio (SNR) in each channel. The performance of the proposed noise reduction algorithm is evaluated with nine Clarion CII cochlear implant patients using IEEE sentences embedded in multi-talker babble and speech-shaped noise at $0-10$ dB SNR. Results indicate that the sigmoidal-shaped weighting function produces significant improvements to speech recognition compared to the subjects' daily strategy. Much of the success of the proposed noise reduction algorithm is attributed to the improved temporal envelope contrast.
© 2007 Acoustical Society of America

1. Introduction

Although many cochlear implant (CI) users are enjoying high levels of speech understanding in quiet environments, noisy listening conditions remain challenging for most. A number of preprocessing noise-reduction algorithms have been proposed for cochlear implants over the years (Yang and Fu, 2005; Loizou *et al.*, 2005; Loizou, 2006; Van Hoesel and Clark, 1995; Wouters and Vanden Berghe, 2001). The preprocessing approach to noise reduction, however, has three main drawbacks: (1) preprocessing algorithms sometimes introduce unwanted distortion in the signal, (2) some algorithms (e.g., subspace algorithms) are computationally complex (and consequently power hungry) and do not integrate well with existing CI strategies, and (3) it is not easy to optimize the operation of a particular algorithm to individual users.

Ideally, noise reduction algorithms should be easy to implement and be integrated into existing coding strategies. In this paper, we propose a simple noise reduction algorithm that can be easily integrated in existing strategies used in commercially available devices. The proposed algorithm fits into the general category of algorithms that perform noise suppression by spectral modification (e.g., spectral subtraction, Wiener filtering—see review in Loizou, 2007). The enhanced envelopes are obtained by applying a weight (taking values in the range of 0 to 1) to the noisy envelopes of each channel. The weights are chosen to be inversely proportional to the estimated SNR of each channel. Envelope amplitudes in channels with high SNR are multiplied by a weight close to one (i.e., left unaltered), while envelope amplitudes in channels with low SNR are multiplied by a weight close to zero (i.e., heavily attenuated). The underlying assumption is that channels with low SNR are heavily masked by noise and therefore contribute little, if any, information about the speech signal. As such, these low-SNR channels are heavily attenuated (or annihilated) leaving only the high-SNR channels, which likely contribute more useful information to the listener.

---

[a]Author to whom correspondence should be addressed: Department of Electrical Engineering, University of Texas at Dallas, P.O. Box 830688, EC 33, Richardson, TX 75083-0688.
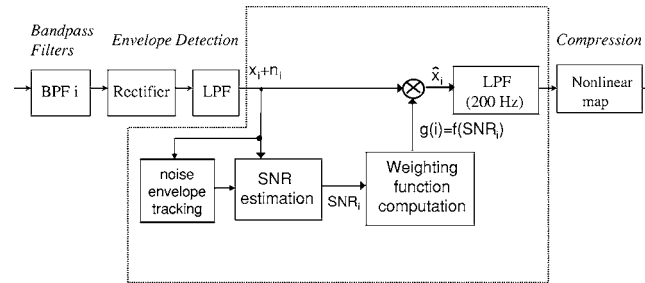
Fig. 1. Block diagram of the proposed noise reduction for the *i*th channel.

2. Experimental design

2.1 Subjects

A total of nine postlingually deafened Clarion CII implant users participated in this experiment. All subjects had at least 3 years of experience with their implant device. Most subjects visited our lab two times. During the first visit, all nine subjects were tested on music perception tasks and other psychophysical tasks unrelated to the current investigation. Due to the limited time available, the subjects were also tested on a single noise condition (5 dB SNR babble). Subjects were subsequently invited again to our lab, but due to various reasons (e.g., health, schedule conflicts), only five of the nine subjects were able to visit our lab to continue the testing for the other noise conditions.

2.2 Noise-suppression algorithm

Figure 1 shows the block diagram of the proposed noise reduction algorithm. The noisy speech signal is bandpass filtered into 16 channels and the envelopes are detected in each channel after full-wave rectification and low-pass filtering (200 Hz, sixth-order Butterworth). The noisy envelopes in each channel are multiplied by channel-specific weighting functions taking values in the range of zero to one depending on the estimated SNR of that channel. The envelopes attenuated by the channel weighting functions are smoothed with a low-pass filter (200 Hz) and log-compressed to the subject's electrical dynamic range. The low-pass filter is used to ensure that the enhanced envelopes are smoothed and are free of any abrupt amplitude changes that may be introduced by the application of the time-varying weighting function.

There are two major components in the proposed algorithm: SNR estimation and computation of the weighting function, which in turn depends on the estimated SNR. These components are discussed next.

2.2.1 Weighting function

We considered using a weighting function that applies heavy attenuation in channels with low SNR and little or no attenuation in channels with high SNR. With that in mind, we chose to use the following sigmoidal-shaped function:

$$g(i,l) = e^{-\beta/\xi(i,l)}, \tag{1}$$

where $\beta = 2$, $g(i,l)$ is the weighting function ($0 < g(i,l) \leq 1$), and $\xi(i,l)$ denotes the *estimated* instantaneous SNR in the *i*th channel at stimulation cycle *l*. This weighting function plateaus at one for SNR > 20 dB and floors to 0 for SNR < -5 dB. The above function was chosen as it has a sigmoidal shape similar to the human listener's psychometric function of intelligibility versus SNR. Other functions with similar shape could alternatively be used. Following the weighting function computation in Eq. (1), the enhanced temporal envelope is obtained by
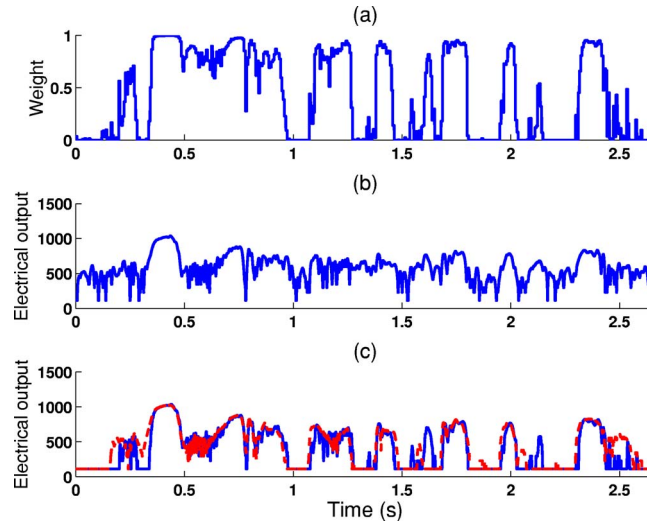
Fig. 2. (Color online) Example plots of attenuation (top panel) applied over time to the noisy envelope (middle panel) to obtain the enhanced envelope (bottom panel) for an IEEE sentence embedded in 5 dB SNR babble. The clean envelope is superimposed (dashed lines) in the bottom panel for comparison.

$$\hat{x}(i,l) = g(i,l) \cdot y(i,l), \tag{2}$$

where $y(i,l)$ is the noisy envelope of the ith channel at cycle l and $\hat{x}(i,l)$ contains the enhanced signal envelope. The enhanced envelope $\hat{x}(i,l)$ is smoothed by the same low-pass filter (200 Hz cutoff frequency) used in the envelope detection, and finally log compressed to produce the electrical amplitudes for stimulation (see Fig. 1). The low-pass filter is used to ensure that the enhanced envelopes are smoothed and are free of any abrupt amplitude changes that may be introduced by the application of the time-varying weighting function.

Figure 2 shows an example plot of the noisy speech envelope in comparison with the enhanced speech envelope for an IEEE sentence ("Every word and phrase he speaks is true.") in 5 dB SNR babble. Only the envelopes for channel 1 (center frequency=398 Hz) are plotted. The top panel (Fig. 2) shows the attenuation values applied to the noisy envelope. For the most part, the attenuation values (i.e., weights) were near one for envelope peaks and close to zero for envelope valleys. The envelope peaks were thus left unaltered while the valleys were attenuated. The resulting enhanced envelope (bottom panel in Fig. 2) had improved temporal envelope contrast. It should be noted that the processed stimuli were not intermittent despite the envelope dips seen in Fig. 2. When a weight of zero is applied to the noisy envelopes, the resulting envelope amplitude is set to the corresponding (electrical) threshold value and not to zero. The dips seen in the enhanced envelopes (Fig. 2) of channel 1 are due to the silence/closures already present in the original clean stimulus (also shown in bottom panel of Fig. 2) or to high-frequency consonants (e.g., /s/).

2.2.2 SNR estimation

The computation of the weighting function in Eq. (1) depends on the estimation of the instantaneous SNR $\xi(i,l)$ of channel i at cycle l, which is obtained using a variant approach reported by Ephraim and Malah (1984):

$$\xi(i,l) = \begin{cases} \alpha \dfrac{\hat{x}^2(i,l-1)}{\hat{n}^2(i,l-1)} + (1-\alpha)\max(\gamma(i,l)-1,0), & i \leq 10(f \leq 1.8 \text{ kHz}), \\ \max(\gamma(i,l)-1,\varepsilon), & i > 10(f > 1.8 \text{ kHz}), \end{cases} \tag{3}$$

where $\hat{x}(i, l-1)$ is the enhanced signal envelope obtained in the last stimulation cycle, ε is a small constant (10^{-6}) needed to avoid possible division by zero in Eq. (1), $\hat{n}(i, l)$ is the estimated envelope amplitude of the noise obtained using a noise-tracking algorithm (Cohen and Berdugo, 2002), α is a smoothing parameter ($0 < \alpha < 1$), and $\gamma(i, l) = y^2(i, l)/\hat{n}^2(i, l)$. The reason for using different methods for estimating the SNR for high ($i > 10$) and low ($i \leq 10$) frequency channels is to allow for faster tracking of sudden changes to the instantaneous SNR in the high frequency channels and relatively slower changes to the instantaneous SNR in the low frequency channels. Unlike Ephraim and Malah (1984) who used a high value for α ($\alpha = 0.98$), we found that a smaller smoothing constant was necessary for faster tracking of the instantaneous SNR, which is computed on a sample-by-sample basis rather than every 20-ms frame. In our study, the smoothing parameter α was set to $\alpha = 0.4$ and $\alpha = 0.6$ based on pilot experimental data. A noise-estimation algorithm (Cohen and Berdugo, 2002) is used to continuously track and update the noise envelope amplitude $\hat{n}(i, l)$ even during speech activity.

2.3 Procedure

The listening task involved sentence recognition in noise. IEEE sentences (IEEE, 1969) corrupted in multi-talker babble (ten female and ten male talkers) and speech-shaped noise were used in the test. Subjects were tested in four different SNR levels: 5 and 10 dB SNR in babble and 0 and 5 dB SNR in speech-shaped noise. Lower SNR levels were chosen for the speech-shaped noise conditions to avoid ceiling effects, as most subjects performed very well at the 10 dB SNR level. The IEEE sentences were recorded in our lab in a double-walled sound-attenuating booth and are available from Loizou (2007). The babble recording was taken from the AUDITEC CD (St. Louis, MO). Two sentence lists (ten sentences per list) were used for each condition. The sentences were processed offline in MATLAB by the proposed algorithm and presented to the subjects using the Clarion CII research platform at a comfortable level. For comparative purposes, subjects were also presented with unprocessed noisy sentences using the experimental processor. Sentences were presented to the listeners in blocks, with 20 sentences per block per condition. Different sets of sentences were used in each condition. Subjects were instructed to write down the words they heard. The presentation order of the processed and control (unprocessed sentences) conditions was randomized among subjects.

3. Results

The sentences were scored in terms of percent of words identified correctly (all words were scored). Figure 3 shows the individual scores for all subjects for the multi-talker babble (5–10 dB SNR) conditions and Fig. 4 shows the individual scores for a subset of the subjects tested in the speech-shaped noise (0–5 dB SNR) conditions. The scores for the subjects (S1, S5, S6, S8, S9) who visited our lab two times were averaged across the two visits (5 dB SNR, Fig. 3, top panel).

ANOVA (with repeated measures) showed a highly significant ($F[2,16] = 14.4, p < 0.0005$) effect of the noise reduction algorithm on speech intelligibility for the 5-dB SNR babble condition. Scores obtained with the proposed noise reduction algorithm were significantly ($p < 0.005$) higher than the scores obtained with the subject's daily processor for both values of α. As shown in Fig. 3, all subjects benefited to some degree with the noise-reduction algorithm. Subjects S7 and S8 in particular received large benefits as their scores nearly doubled. With the exception of subjects S3, S5, S6, and S9, subjects performed equally well with $\alpha = 0.4$ and $\alpha = 0.6$ in Eq. (3). ANOVA (with repeated measures) performed on the 10-dB SNR babble data also showed a highly significant ($F[2,8] = 24.3, p < 0.0005$) effect of the noise reduction algorithm on speech intelligibility. The mean improvement in performance with the proposed noise reduction algorithm was substantially larger than that obtained in the 5-dB SNR condition. Mean scores improved from 47% correct to 71% correct, with a small variability among subjects. The large benefit in intelligibility was consistent for all five subjects.
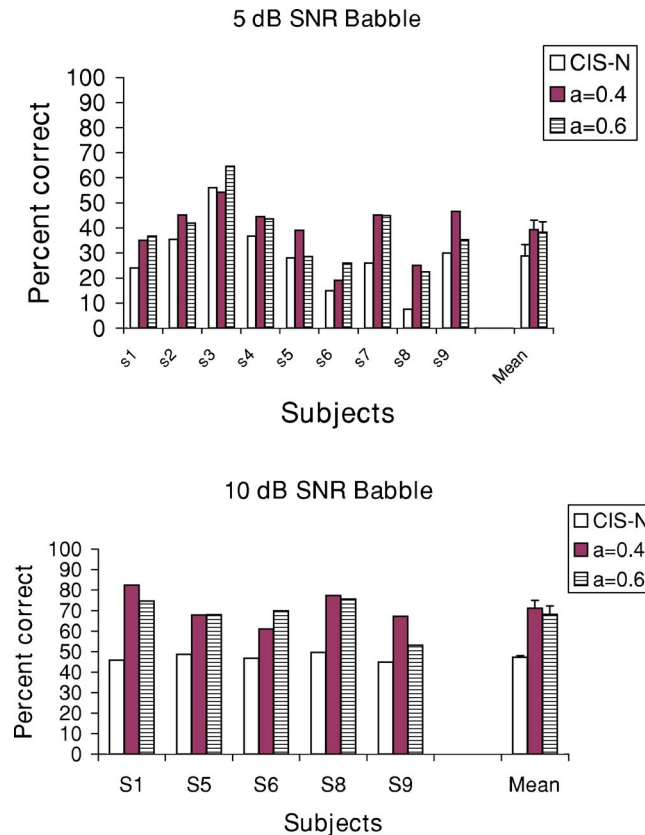
Fig. 3. (Color online) Subjects' performance on identification of words in sentences embedded in 5 and 10 dB SNR multi-talker babble and processed by the proposed algorithm with $\alpha=0.4$ and $\alpha=0.6$. Subjects' baseline performance (CIS-N) on unprocessed noisy stimuli is indicated by white bars. Error bars indicate standard errors of the mean.

ANOVA (with repeated measures) performed on the 0-dB SNR speech-shaped noise data showed a highly significant ($F[2,8]=9.5, p=0.008$) effect of the noise reduction algorithm on speech intelligibility. Particularly large improvements in performance were noted for subjects S1, S8, and S9. ANOVA (with repeated measures) performed on the 5-dB SNR speech-shaped noise data did not show a significant effect ($F[2,8]=0.9, p=0.445$). Performance obtained with the proposed noise reduction algorithm was as good as that obtained with the subject's daily processor.

4. Discussion

The above analysis clearly indicates that the proposed sigmoidal-shaped function provided significant benefits to CI users in nearly all conditions. We believe that much of the success of the proposed noise reduction algorithm can be attributed to the improved temporal envelope contrast. As shown in the example in Fig. 2, the sigmoid function preserves the envelope peaks and deepens the envelope valleys, thereby increasing the effective envelope dynamic range within each channel. Note, for instance, in Fig. 2, the change in the noisy envelope at $t=2.3$ s corresponding to the phoneme /s/ in the latter portion of the sentence. The depicted noisy envelope of channel 1 (center frequency=382 Hz) is near 50% of the electrical dynamic range when, in fact, it should have been near threshold level, since much of the energy of /s/ is concentrated in the high-frequency channels. After applying the sigmoidal-shaped function, the noisy envelope was
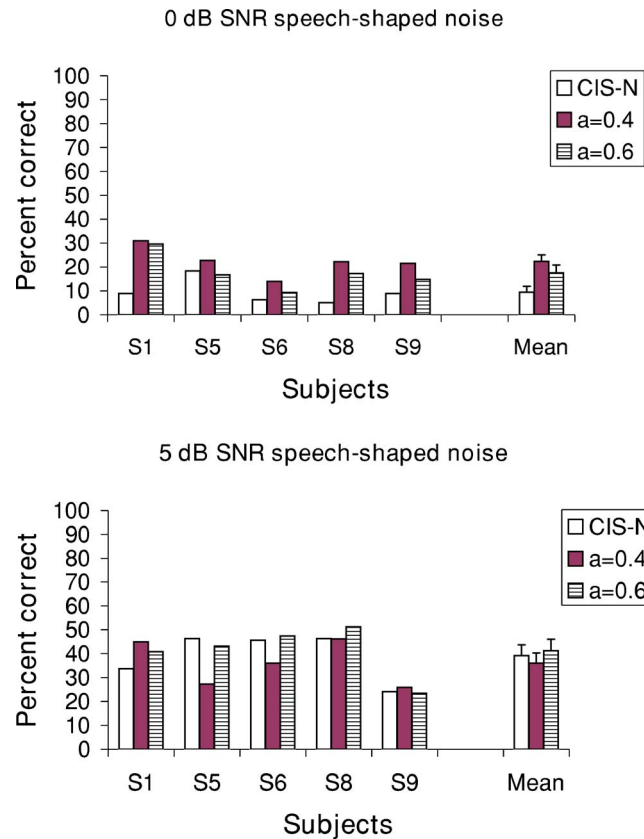
Fig. 4. (Color online) Subjects' performance on identification of words in sentences embedded in 0 and 5 dB SNR speech-shaped noise and processed by the proposed algorithm with $\alpha=0.4$ and $\alpha=0.6$. Subjects' baseline performance (CIS-N) on unprocessed noisy stimuli is indicated by white bars. Error bars indicate standard errors of the mean.

reduced to near threshold level. In brief, the proposed weighting function can make the envelope peaks and envelope valleys more discernible and perhaps more accessible to the CI users (compare the middle and bottom panels in Fig. 2).

    The attenuation applied to the noisy envelopes is computed inversely proportional to the estimated SNR of each channel and is applied to both the speech and noise envelope signals. Clearly the attenuation itself cannot eliminate the noise as it is applied to both speech and noise signals. Yet, the applied attenuation, estimated according to the sigmoidal-shaped function, improves speech intelligibility for two main reasons. First, it is applied selectively and inversely proportional to the estimated SNR of each channel. More attenuation is applied to low-SNR channels where the speech signal is heavily masked by the noise, and is probably unintelligible. These channels provide unreliable and perhaps distracting information to the CI users and should therefore be attenuated. In contrast, no attenuation is applied to the channels in which the SNR is sufficiently high, thus enabling the listeners to ignore the masker. The high-SNR channels carry perhaps the most reliable information about the underlying signal and are therefore left unaltered. From Fig. 2 we observe that the SNR estimation does not need to be very accurate in terms of computing the exact weight to be applied to the noisy envelopes. It suffices if the SNR estimation algorithm performs sufficiently well in terms of discriminating high from low SNR envelopes, since the assigned weight will be either near the value of one (for SNR >20 dB) or near the value of zero (for SNR <-5 dB), respectively. Secondly, the temporal envelope contrast is improved (see example in Fig. 2). As a result, the envelope peaks and

valleys are more discernible to the CI users who have a limited electrical dynamic range. This in turn makes the consonant/vowel boundaries clearer and more accessible to the CI users.

Overall, our algorithm compares favorably against other single-microphone methods proposed for cochlear implants (e.g., Yang and Fu, 2005; Loizou *et al.* 2005). A larger improvement (10–25 percentage points) in performance was obtained with our proposed method in multi-talker babble compared to the improvement (7 percentage points) obtained by the preprocessing method reported in Yang and Fu (2005). Other advantages of the proposed method include the lack of algorithmic delay associated with preprocessing techniques, low computational complexity, and ease of integration in existing CI strategies.

Acknowledgments

References and links

Cohen, I., and Berdugo, B. (**2002**). "Noise estimation by minima controlled recursive averaging for robust speech enhancement," IEEE Signal Process. Lett. **9**, 12–15.

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**(6), 1109–1121.

IEEE. (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **AU-17(3)**, 225–246.

Loizou, P. (**2006**). "Speech processing in vocoder-centric cochlear implants," Adv. Oto-Rhino-Laryngol. **64**, 109–143.

Loizou, P. (**2007**). *Speech Enhancement: Theory and Practice* (CRC Press, Boca Raton, FL).

Loizou, P., Lobo, A., and Hu, Y. (**2005**). "Subspace algorithms for noise reduction in cochlear implants," J. Acoust. Soc. Am. **118**(5), 2791–2793.

Van Hoesel, R., and Clark, G. (**1995**). "Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients," J. Acoust. Soc. Am. **97**(4), 2498–2503.

Wouters, J., and Vanden Berghe, J. (**2001**). "Speech recognition in noise for cochlear implantees with a two-microphone monaural adaptive noise reduction system," Ear Hear. **22**(5), 420–430.

Yang, L., and Fu, Q. (**2005**). "Spectral subtraction-based speech enhancement for cochlear implant patients in background noise," J. Acoust. Soc. Am. **117**(3), 1001–1004.

Speech disruption during delayed auditory feedback with simultaneous visual feedback

Jeffery A. Jones

Centre for Cognitive Neuroscience and Department of Psychology, Wilfrid Laurier University, Waterloo, Ontario N2L 2C5, Canada
jjones@wlu.ca

Danielle Striemer

Department of Psychology, Wilfrid Laurier University, Waterloo, Ontario N2L 2C5, Canada
dstriemer@gmail.com

Abstract: Delayed auditory feedback (DAF) regarding speech can cause dysfluency. The purpose of this study was to explore whether providing visual feedback in addition to DAF would ameliorate speech disruption. Speakers repeated sentences and heard their auditory feedback delayed with and without simultaneous visual feedback. DAF led to increased sentence durations and an increased number of speech disruptions. Although visual feedback did not reduce DAF effects on duration, a promising but nonsignificant trend was observed for fewer speech disruptions when visual feedback was provided. This trend was significant in speakers who were overall less affected by DAF. The results suggest the possibility that speakers strategically use alternative sources of feedback.

1. Introduction

Numerous laboratory studies have shown that altering the auditory feedback speakers hear affects ongoing speech production. For example, exposing speakers to increases in environmental noise causes speakers to increase their speaking volume and the duration of their utterances (Lane and Tranel, 1971; Bauer *et al.*, 2006). Selectively filtering frequencies (Garber and Moller, 1979) or modifying the fundamental frequency (Elman, 1981; Burnett *et al.*, 1997; Kawahara, 1998; Jones and Munhall, 2000) or the formant frequencies (Houde and Jordan, 1998; Purcell and Munhall, 2006) elicits compensatory productions that mitigate the alterations. However, the most profound disruptions to ongoing vocal productions result when speakers hear their ongoing speech delayed (Lee, 1950). Exposure to delayed auditory feedback (DAF) often results in "stutterlike" disturbances in fluency (Fairbanks, 1955; Fairbanks and Guttman, 1958). These disturbances include speaking-rate decreases, increased speech intensity and pitch, syllable repetitions and omissions, and misarticulation (Black, 1951; Atkinson, 1953; Yates, 1963; Howell and Archer, 1984).

The profound effects caused by DAF (and the effects of certain other altered feedback conditions) led to speculation that speech production is monitored in a closed-loop manner (Lee, 1950; Fairbanks, 1954). According to these servomechanistic accounts, a comparator looks for discrepancies between the intended output of a vocal production and the sensory feedback; the disruptions observed during DAF are a manifestation of the corrective action initiated to overcome the perceived mismatch. However, Borden (1979) contended that speech rate is too quick for auditory feedback to be processed and the corrections implemented before the next segment is produced. Moreover, Howell and Archer (1984) showed that when a 500-Hz square wave matching the amplitude envelope of the speaker's speech was substituted for the delayed speech signal, speakers suffered a similar reduction in their speech rate, as they did when they heard their true speech signal delayed. Indeed, DAF effects are not limited to speech behaviors and are found with other motor behaviors such as tapping and music production

(Chase *et al.*, 1959; Smith *et al.*, 1960; Finney and Warren, 2002). Thus, the root cause of DAF effects appears to be due to a more general disruption of the temporal relationship between production and acoustic input and not due to the fact that the feedback system receives incorrect information about the specific articulations (Howell and Sackin, 2002).

There has been limited research conducted on the remedial effects of providing alternative forms of synchronous feedback simultaneously with DAF. Howell and Archer (1984) reported that increasing the volume of DAF led to increased levels of disruption. This finding may suggest that speakers can use their veridical feedback to reduce DAF effects (either transmitted through bone or air). However, given that DAF effects are likely the result of the detection of global asynchronies between production and feedback, it is probably not necessary that an alternative source of feedback provide more than a crude indication of synchrony. Another naturally synchronous signal during speech is the visible movements of the speaker's face. A wealth of research has shown that listeners readily use visual speech cues: whether in noisy environments or under optimal listening conditions, information from a speaker's face significantly enhances auditory intelligibility (Sumby and Pollack, 1954; Davis and Kim, 2004). The current study investigates whether providing speakers with visual feedback regarding the timing of their ongoing speech can reduce DAF effects.

An earlier study conducted by Tye-Murray (1986) found no evidence that visual information could be used to improve speech production. In her study, Tye-Murray asked 11 volunteers to say eight sentences while hearing their voice delayed, with and without the availability of a mirror to monitor their productions. However, Tye-Murray only examined sentence duration and did not look at the number of speech disruptions that occurred. The results showing that the duration of sentence repetition was unaffected by the presence of the visual feedback do not completely exclude the possibility that the number of speech disruptions was reduced. In the current study speakers heard their auditory feedback delayed by 180 ms while they produced sentences with and without visual feedback. Any moderating effect of providing visual feedback was evaluated by measuring both sentence duration and the number of speech disruptions. We additionally looked at whether the provision of visual feedback would differentially affect those speakers who were more affected by DAF compared to those speakers who were less affected. One might predict that if participants experience very few speech disruptions under DAF, it is unlikely that the availability of visual feedback will further reduce the small number of speech disruptions that occur. However, individuals who experience a greater degree of disruption under DAF may benefit more from the availability of visual feedback. Alternatively, it could be the case that individuals who experience fewer speech disruptions under DAF are *better* able to integrate alternative sources of sensory information to aid in speech production. Conversely, individuals who experience greater speech disruption may be *less* able to integrate other sources of sensory information to aid their speech production.

2. Methods

2.1 Participants

Twenty-two right-handed men (mean age 21.6 years) participated in this study. However, data from only 20 participants were analyzed because two of them were statistical outliers (i.e., their data values were over three times the interquartile range, above the third quartile). All 20 remaining participants were university students with no reported prior neurological damage, or speech or language disorders. All participants had normal or corrected-to-normal vision. The procedures were approved by the Wilfrid Laurier University Research Ethics Board and all participants gave informed consent.

2.2 Apparatus and procedure

Participants sat in a double-walled sound booth and wore headphones (Sennheiser HD 280 Pro) and a headset microphone (AKG C420) during the experimental session. Participants' vocal productions were recorded as they repeated the same ten sentences in each experimental condition. Each sentence was made up of between five to eight words (eight to ten syllables). To

familiarize participants with the stimuli, they were asked to read the sentences aloud before commencing the experiment. On each trial, participants heard a recording of a sentence and were asked to repeat the sentence at a consistent pace.

Each speaker participated in the six following conditions: (1) The "NAF" (normal auditory feedback) condition was a control condition and served as a baseline for the other five experimental conditions. In this condition, participants repeated the ten sentences while receiving unaltered auditory feedback. They simultaneously stared at a black fixation cross in a 6 $\times$ 6 in. white square over a blue background on a 17-in. monitor. (2) The "DAF" condition required participants to stare at the fixation cross on the screen and repeat the same sentences while their auditory feedback was delayed by 180 ms using a digital signal processor (Tucker-Davis Technologies, RX6 Multifunction Processor). (3) The "NAF mirror" condition required participants to repeat the sentences while hearing their auditory feedback presented without delay, while viewing the movements of their face in a mirror measuring 7 in. in diameter at an approximately constant distance. The mirror and microphone were adjusted so the participants could view their mouth movements as the sentences were recited. (4) The "DAF mirror" condition required participants to repeat the sentences while their auditory feedback was delayed and they simultaneously received visual feedback regarding their mouth movements reflected by the mirror. (5) The "NAF sentence" condition required participants to repeat the sentences while reading the sentences presented orthographically on the 17-in. monitor and hearing their auditory feedback presented without delay. (6) The "DAF sentence" condition required participants to repeat the sentences while their auditory feedback was delayed and they read the sentences on the monitor. Comparing performance during this condition to performance during the other DAF conditions served to ensure that any deficits observed did not result from misremembering the sentences.

Sentence order was randomized across conditions and the order of conditions was counterbalanced across participants. Gaussian noise was presented throughout the experimental session in an effort to mask the participants' real time auditory feedback.

2.3 Data analysis

The durations of utterances were determined manually using Praat (Boersma, 2001). Speech disruptions were identified as a word or syllable repetition, part-word prolongation, inaudible postural fixation, or misarticulation of a word. Both sentence duration and number of speech disruptions were analyzed separately using a 2×3 repeated measures ANOVA with auditory feedback (normal, delayed) and visual cue (fixation, mirror, sentence) as the within subject factors. To determine whether participants who were more affected by DAF responded differently to the presence of visual feedback compared to those participants who were less affected, we performed a median split based on the number of speech disruptions that occurred for each participant during the "DAF" baseline condition. Two groups were formed: a "low"-disruption group ($n=10$) and a "high"-disruption group ($n=10$). The number of speech disruptions made by these two groups was analyzed using the same 2×3 repeated measures ANOVA described above.

3. Results

Two participants were identified as statistical outliers and thus were not included in the analysis. Data from the remaining 20 participants were analyzed. Results of the statistical analysis of sentence duration revealed a main effect of auditory feedback [$F(1,19)=39.54, p<0.001$]; sentence durations in the DAF conditions were longer than sentence durations in the NAF conditions [see Fig. 1(a)]. No other significant effects were observed for sentence duration. The results for the number of speech disruptions also revealed a main effect of auditory feedback [$F(1,19)=23.80, p<0.001$] with a greater number of disruptions made in the DAF conditions than in the NAF conditions [see Fig. 1(b)]. Although there was not a significant interaction
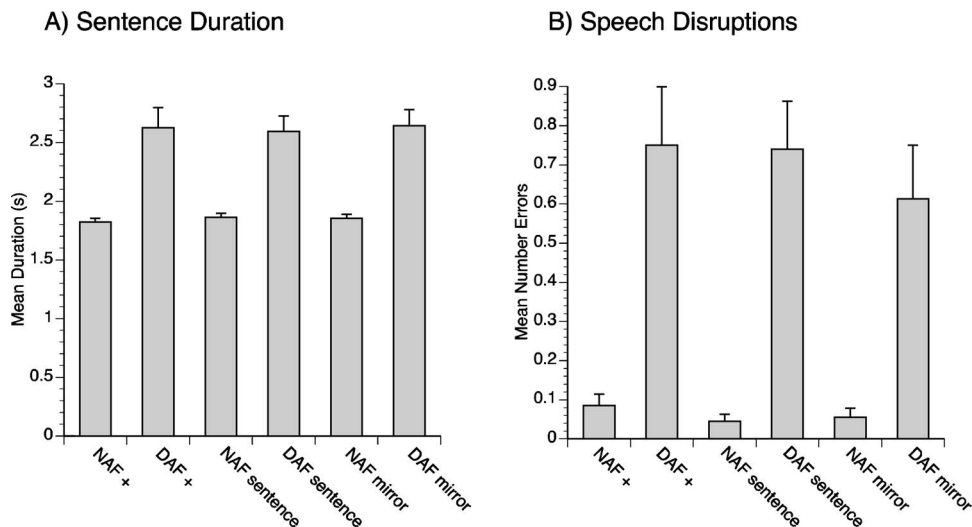
Fig. 1. (a) Mean sentence duration for each of the six experimental conditions. (b) Mean number of speech disruptions for each of the six experimental conditions.

between auditory feedback condition and visual cue condition $[F(1,19)=2.099, p=0.137]$ there was a visible trend for a higher number of speech disruptions to occur during the DAF compared to the DAF with mirror (facial) feedback conditions.

A median split was performed by dividing the participants into a high-disruption group ($n=10$) and a low-disruption group ($n=10$) based on the median number of speech disruptions experienced during the DAF baseline condition. The statistical analysis of the high-disruption group revealed a main effect of auditory stimulus $[F(1,9)=38.00, p<0.001]$ with DAF conditions eliciting a greater number of speech disruptions than the NAF conditions [see Fig. 2(a)]. No other effects were significant in the high-disruption group. For the low-disruption group, Mauchly's test revealed that there was a violation of sphericity $[W(2)=0.47, p=0.048]$, so a multivariate analysis was used. The multivariate analysis revealed a significant main effect
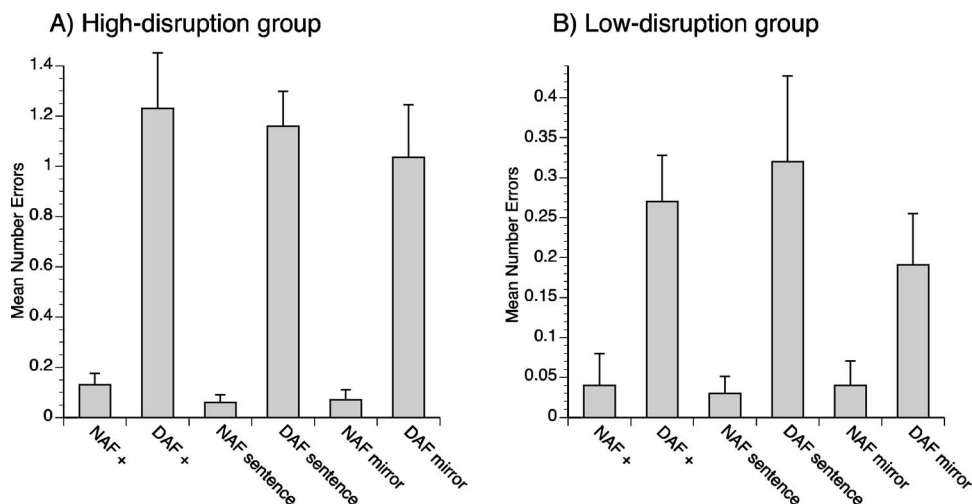
Fig. 2. Mean number of speech disruptions for each of the six experimental conditions in the (a) "high-disruption" and (b) "low-disruption" groups.

of auditory stimulus $[F(1,9)=10.33, p=0.011]$ with DAF conditions leading to more speech disruptions than the NAF conditions [see Fig. 2(b)]. In addition, there was a significant interaction between auditory stimulus and visual stimulus $[F(2,8)=4.67, p=0.045]$. Posthoc paired samples t tests revealed significantly fewer speech disruptions occurred during the DAF mirror condition compared to the DAF sentence condition $[t(9)=2.251, p=0.05]$. No other significant differences were found.

4. Discussion

In the present study, sentence duration and number of speech disruptions were examined in order to determine whether visual cues could reduce the speech disruption typically observed with exposure to DAF. Delaying auditory feedback did consistently lead to increased sentence duration, but providing visual feedback simultaneously with DAF did not reduce this effect. Moreover, all DAF conditions led to a larger number of speech disruptions than when real time auditory feedback was available. In addition, although not reliable, on average a lower disruption rate was observed when speech was delayed and visual feedback was provided compared to when speakers fixated on a fixation point on a computer monitor or read the sentence on the screen. This trend was also apparent when we split our participants into a high-disruption group and a low-disruption group. However, the trend was only statistically significant for the low-disruption group who experienced fewer disruptions when they viewed their face with DAF than when they read the sentences.

   Our results both replicate and extend those observed by Tye-Murray (1986). Tye-Murray also found that providing speakers with visual feedback regarding their speech production under conditions with DAF did not reduce durational effects. Based on those finding, she concluded that visual feedback does not diminish the effects of DAF. However, we found promising trends that suggest that visual feedback might instead reduce the number of speech disruptions that speakers experience under DAF.

   It is well known that listeners readily use visual speech cues (i.e., facial movements) during speech perception (e.g., Davis and Kim, 2004). In fact, brain imaging work demonstrates that even in the absence of audible speech, visual speech cues activate the supratemporal auditory cortex in normal-hearing individuals (Calvert *et al.*, 1997). However, we believe that it is unlikely that DAF effects arise from a close monitoring of the precision of production. Rather, we believe that DAF effects arise from monitoring a signal that is globally asynchronous (Howell and Sackin, 2002). The visual feedback we provided to speakers was inherently synchronous with their ongoing production and speakers may have been able to monitor this alternative feedback to some degree, thereby reducing the speech disruption caused by exposure to DAF.

   Although consistent, this reduced speech disruption was statistically unreliable except for those speakers who performed best under the DAF condition. The reason for this is unclear, but it is known that speakers do adopt strategies to overcome DAF effects (Katz and Lackner, 1977). One possible strategy that speakers may adopt is to ignore the auditory feedback they receive and use alternative feedback pathways. Thus, it is possible that those speakers who experienced fewer speech disruptions were better able to ignore their auditory feedback and focus on alternative feedback sources that were synchronous with their production (e.g., proprioception). The addition of the visual feedback may have enhanced their ability to ignore the DAF and use visual feedback as an alternative and congruent feedback source. It is possible that both high and low performing speakers could be trained to take advantage of visual feedback. Indeed, previous research has demonstrated that DAF effects can be reduced with practice (Goldiamond *et al.*, 1962). Perhaps prolonged practice with visual feedback could lead to more pronounced reductions in the disruption rates we observed.

   The possibility that visual feedback can ameliorate some of the effects caused by DAF has implications for understanding the remedial effects of similar treatments for certain speech disorders. For example, Lee (1950) was among the first to note the similarity of DAF effects to stuttering and since then a number of researchers have hypothesized the irregular use of auditory feedback as a cause of stuttering (Fairbanks, 1954; 1955; Max *et al.*, 2004). This view was

paradoxically bolstered by reports that dysfluency could be reduced by providing people who stutter with DAF (Kalinowski *et al.*, 1993; 1996; Van Borsel *et al.*, 2003; Alm, 2006). How altered auditory feedback reduces stuttering is the subject of intense debate. However, our results are especially interesting in light of Kalinowski and colleague's recent report that fluency is enhanced when people who stutter view the silent articulations of another speaker producing the same utterance as themselves (Kalinowski *et al.*, 2000).

In closing, our data suggest speakers may be able to use visual feedback to reduce the speech disruption that results from exposure to DAF. The literature on DAF has in large part focused on the strategies that speakers use to regain fluency under DAF conditions. An accepted assumption of some researchers is that speakers learn to ignore auditory feedback and may avail themselves of alternative feedback sources such as proprioception. These strategies may in fact underlie the beneficial effects observed in clinical populations where DAF is used as a treatment for stuttering. However, it is difficult to test these assumptions because most obvious forms of alternative feedback are difficult or impossible to manipulate experimentally. Visual feedback is easily manipulated and offers a window into speakers' strategic use of alternative sources of feedback. It will be interesting in future work to manipulate factors such as the salience of the visual feedback relative to the auditory feedback, as well as the amount of practice and training speakers receive.

Acknowledgments

References and links

Alm, P. A. (**2006**). "Stuttering and sensory gating: A study of acoustic startle prepulse inhibition," Brain Lang **97**, 317–321.

Atkinson, C. J. (**1953**). "Adaptation to delayed sidetone," J. Speech Hear Disord. **18**, 386–391.

Bauer, J. J., Mittal, J., Larson, C. R., and Hain, T. C. (**2006**). "Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude," J. Acoust. Soc. Am. **119**, 2363–2371.

Black, J. W. (**1951**). "The effect of delayed sidetone upon vocal rate and intensity," J. Speech Hear Disord. **16**, 56–60.

Boersma, P. (**2001**). "Praat, a system for doing phonetics by computer," Glot Int. **5**, 341–345.

Borden, G. J. (**1979**). "An interpretation of research on feedback interruption in speech," Brain Lang **7**, 307–319.

Burnett, T. A., Senner, J. E., and Larson, C. R. (**1997**). "Voice F0 responses to pitch-shifted auditory feedback: A preliminary study," J. Voice **11**, 202–211.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodruff, P. W., Iversen, S. D., and David, A. S. (**1997**). "Activation of auditory cortex during silent lipreading," Science **276**, 593–596.

Chase, R. A., Harvey, S., Standfast, S., Rapin, I., and Sutton, S. (**1959**). "Comparison of the effects of delayed auditory feedback on speech and key tapping," Science **129**, 903–904.

Davis, C., and Kim, J. (**2004**). "Audio-visual interactions with intact clearly audible speech," Q. J. Exp. Psychol. A **57**, 1103–1121.

Elman, J. L. (**1981**). "Effects of frequency-shifted feedback on the pitch of vocal productions," J. Acoust. Soc. Am. **70**, 45–50.

Fairbanks, G. (**1954**). "Systematic research in experimental phonetics. I. A theory of the speech mechanism as a servosystem," J. Speech Hear Disord. **19**, 133–139.

Fairbanks, G. (**1955**). "Selective vocal effects of delayed auditory feedback," J. Speech Hear Disord. **20**, 333–345.

Fairbanks, G., and Guttman, N. (**1958**). "Effects of delayed auditory feedback upon articulation," J. Speech Hear. Res. **1**, 12–22.

Finney, S. A., and Warren, W. H. (**2002**). "Delayed auditory feedback and rhythmic tapping: Evidence for a critical interval shift," Percept. Psychophys. **64**, 896–908.

Garber, S. R., and Moller, K. T. (**1979**). "The effects of feedback filtering on nasalization in normal and hypernasal speakers," J. Speech Hear. Res. **22**, 321–333.

Goldiamond, I., Atkinson, C. J., and Bilger, R. C. (**1962**). "Stabilization of behavior and prolonged exposure to delayed auditory feedback," Science **135**, 437–438.

Houde, J. F., and Jordan, M. I. (**1998**). "Sensorimotor adaptation in speech production," Science **279**, 1213–1216.

Howell, P., and Archer, A. (**1984**). "Susceptibility to the effects of delayed auditory feedback," Percept. Psychophys. **36**, 296–302.

Howell, P., and Sackin, S. (**2002**). "Timing interference to speech in altered listening conditions," J. Acoust. Soc. Am. **111**, 2842–2852.

Jones, J. A., and Munhall, K. G. (**2000**). "Perceptual calibration of FO production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**, 1246–1251.

Kalinowski, J., Armson, J., Roland-Mieszkowski, M., Stuart, A., and Gracco, V. L. (**1993**). "Effects of alterations in auditory feedback and speech rate on stuttering frequency," Lang Speech **36**, 1–16.

Kalinowski, J., Stuart, A., Rastatter, M. P., Snyder, G., and Dayalu, V. (**2000**). "Inducement of fluent speech in persons who stutter via visual choral speech," Neurosci. Lett. **281**, 198–200.

Kalinowski, J., Stuart, A., Sark, S., and Armson, J. (**1996**). "Stuttering amelioration at various auditory feedback delays and speech rates," Eur. J. Disord. Commun. **31**, 259–269.

Katz, D. I., and Lackner, J. R. (**1977**). "Adaptation to delayed auditory feedback," Percept. Psychophys. **22**, 476–486.

Kawahara, H. (**1998**). "Hearing voice: Transformed auditory feedback effects on voice pitch control," in *Computational Auditory Scene Analysis*, edited by D. F. Rosenthal and H. G. Okuno (Erlbaum, Mahwah, NJ), pp. 335–349.

Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.

Lee, B. S. (**1950**). "Some effects of side-tone delay," J. Acoust. Soc. Am. **22**, 639–640.

Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., and Wallace, M. E. (**2004**). "Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering," Contemporary Issues in Communication Science and Disorders **31**, 105–122.

Purcell, D., and Munhall, K. (**2006**). "Compensation following real-time manipulation of formants in isolated vowels," J. Acoust. Soc. Am. **119**, 2288–2297.

Smith, W. M., McCrary, J. W., and Smith, K. U. (**1960**). "Delayed visual feedback and behavior," Science **132**, 1013–1014.

Sumby, W. H., and Pollack, I. (**1954**). "Visual contribution to speech intelligibility in noise," J. Acoust. Soc. Am. **26**, 212–215.

Tye-Murray, N. (**1986**). "Visual feedback during speech production," J. Acoust. Soc. Am. **79**, 1169–1171.

Van Borsel, J., Reunes, G., and Van den Bergh, N. (**2003**). "Delayed auditory feedback in the treatment of stuttering: Clients as consumers," Int. J. Lang Commun. Disord. **38**, 119–129.

Yates, A. J. (**1963**). "Delayed auditory feedback," Psychol. Bull. **60**, 213–232.

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 and notices are 2 months prior to publication.

President's report on the 153rd meeting of the Acoustical Society of America held in Salt Lake City, Utah

The 153rd meeting of the Acoustical Society of America was held 4–8 June at the Hilton Salt Lake City Hotel in Salt Lake City, UT. This is the third time that the Society has met in this city, previous meetings being held in 1979 and 1992.

The meeting drew a total of 924 registrants, including 125 nonmembers, 219 students, and 104 registrants from outside North America. There were 31 registrants from the United Kingdom, 13 from France, 10 from Germany, 6 from Japan, 5 each from Denmark, The Netherlands, and the People's Republic of China, 4 each from Ireland and Taiwan, 2 each from Croatia, India, Portugal, Singapore, and Spain, and 1 each from Australia, Belgium, Chile, Finland, Hong Kong, Italy, Korea, New Zealand, Netherlands Antilles, Oman, and Sweden. North American countries, Canada, Mexico, and the United States, accounted for 16, 2, and 802, respectively.

A total of 683 papers, organized into 84 sessions, covered the areas of interest of all 13 Technical Committees. The meeting also included 18 meetings dealing with standards. The Monday evening tutorial lecture series was continued by Uwe Hansen, Indiana State University, and Jerry Floor, Director of the Salt Lake City Jazz Orchestra. The tutorial "Musical Acoustics: Science and Performance–An Evening with the Salt Lake City Jazz Orchestra" was presented to an audience of about 300 at the Capitol Theatre. This event was attended by both ASA members as well as members of the public.

The Society's 13 Technical Committees held open meetings where they made plans for special sessions at future ASA meetings, discussed topics of interest to the attendees, and held informal socials after the end of the official business. These are working, collegial meetings and all people attending Society meetings are encouraged to attend and to participate in the discussions. More information about Technical Committees, including minutes of meetings, can be found on the ASA website ⟨http://asa.aip.org/committees.html⟩ and in the Acoustical News USA section of JASA in the September, October and November issues.

The ASA Student Council hosted a Student Reception with over 125 people in attendance. This reception, which was subsidized by the National Council of Acoustical Consultants, enabled students to meet with established members of the Acoustical Society of America. Several of the Technical Committees awarded Best Student Paper Awards or Young Presenter Awards to students and young professionals who presented papers at the meeting. The list of award recipients, as well as other information for students, can be found online at the ASA Student Zone website ⟨http://www.acosoc.org/student/⟩.

The Technical Committee on Signal Processing in Acoustics sponsored its eighth Gallery of Acoustics at the Salt Lake City meeting. The winning entry titled "Acoustic Demonstrations of Vowel Production Using Vocal-Tract Models was submitted by Takayuki Arai, Sophia University, Tokyo, who received the $350 prize.

The Technical Committee on Architectural Acoustics sponsored a Student Design Competition, which involved the design of a drama theater complex located within an urban mixed-use development. The entries were judged by a panel of architects and acoustical consultants. The first prize winner received a cash award of $1000 and entries selected for "commendation" awards received $500 for each entry. Award winners are: First Honors—Andrew Miller, Megan Hunziker, and Matt Pauly of the University of Kansas; Commendations—Marshall Hilton, Pamela Harght, and Andrew Thomas, University of Kansas; Dorea Ruggles, Elizabeth Panzer, and Nicholas Tranby, Rensselaer Polytechnic Institute; Tom LePage, Yun Jing, and Xiaohu Chen, Rensselaer Polytechnic Institute; Arthur van der Harten, Joshua Fialkoff, and Rolando de la Cruz, Rensselaer Polytechnic Institute; and Benjamin Jados and Ayse Bautz, Illinois Institute of Technology.

The Salt Lake City meeting committee arranged two technical tours, which attracted enthusiastic participation by meeting attendees. The first was a tour of the acoustics research facilities at Brigham Young University. The second was a tour of the Latter Day Saints (LDS) Conference Center and the Mormon Tabernacle which included the 21 000 seat auditorium that is used for both musical performances and the spoken word and the newly renovated Mormon Tabernacle. At the conclusion of the technical tour, participants were invited to attend an organ recital in the Mormon Tabernacle.

Social events included the two social hours held on Tuesday and Thursday, an "icebreaker" and a reception for students, the Fellows Luncheon, and the morning coffee breaks. A special program for students to meet one-on-one with members of the ASA over lunch, which is held at each meeting, was organized by the Committee on Education in Acoustics. These social events provided the settings for participants to meet in relaxed settings to encourage social exchange and informal discussions. The Women in Acoustics Luncheon was held on Wednesday afternoon and was attended by over 100 people.

Dr. Richard T. Kouzes, Laboratory Fellow at the Department of Energy's Pacific Northwest National Laboratory, was the speaker at the Fellows luncheon, which was attended by about 100 people. The subject of his talk was "Detection of Nuclear Threats at Borders." The Fellows luncheon is open to all meeting attendees.

The Salt Lake City organizing committee also arranged two tours for accompanying persons, which included a tour of Salt Lake City and a tour of Provo Canyon, Robert Redford's Sundance Resort—home of the Sundance Film Institute, the scenic Wasatch mountains, and Park City.

The plenary session included a business meeting of the Society, announcements, acknowledgment of the members and other volunteers who organized the meeting, and the presentation of awards and certificates to newly elected Fellows.

The 2006 Rossing Prize in Acoustics Education was presented to William J. Strong, Brigham Young University (see Fig. 1). Dr. Strong presented the Acoustics Education Prize Lecture titled "Descriptive acoustics of music and speech" earlier in the meeting.

FIG. 1. (Color online) ASA President Anthony Atchley (l) presents the 2006 Rossing Prize in Acoustics Education to William J. Strong (r). Photo by Kenny Crookston.

FIG. 2. (Color online) ASA President Anthony Atchley (r) presents 2007 Medwin Prize in Acoustical Oceanography to Brian Dushaw (l). Photo by Kenny Crookston.

The 2007 Medwin Prize in Acoustical Oceanography was presented to Brian Dushaw, University of Washington, "for contributions to acoustic measurement of climate variability" (see Fig. 2). Dr. Dushaw presented the Acoustical Oceanography Prize Lecture titled "The recent history of our understanding of low-mode internal tides in the ocean" earlier in the meeting.

The R. Bruce Lindsay Award was presented to Dorian S. Houser, BIO-MIMETICA, "for contributions to animal bioacoustics and to understanding echolocation and hearing by dolphins" (see Fig. 3). The Helmholtz-Rayleigh Interdisciplinary Silver Medal in Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics was presented to Edwin L. Carstensen, University of Rochester, "for contributions to the physics of biomedical ultrasound" (see Fig. 4). The Gold Medal was presented to Katherine S. Harris,

FIG. 4. (Color online) ASA President Anthony Atchley (l) presents the 2007 Helmholtz-Rayleigh Interdisciplinary Silver Medal to Edwin Carstensen (r). Photo by Kenny Crookston.

City University of New York Graduate Program (ret.), "for pioneering research and leadership in speech production and dedicated service to the Society." (see Fig. 5).

Election of eleven members to Fellow grade was announced and fellowship certificates were presented. New fellows are: Michael R. Bailey, Paul E. Barbone, Damian J. Doria, John M. Eargle, Stephen J. Elliott, Hedwig E. Gockel, Kai Ming Li, T. Douglas Mast, Albert Migliori, Claire Prada, and Karim Sabra (see Fig. 6).

ASA President Anthony Atchley expressed the Society's thanks to the Local Committee for the excellent execution of the meeting, which clearly evidenced meticulous planning. He introduced Scott D. Sommerfelt (see Fig. 7), Chair of the Salt Lake City meeting, who acknowledged the contributions of the members of his committee including: Timothy W. Leishman, Technical Program Chair; Tracianne B. Neilsen, Food Service/Social

FIG. 3. (Color online) ASA President Anthony Atchley (l) presents the 2007 R. Bruce Lindsay Award to Dorian S. Houser (r). Photo by Kenny Crookston.

FIG. 5. (Color online) ASA President Anthony Atchley (r) presents the Gold Medal to Katherine Harris (l). Photo by Kenny Crookston.

FIG. 6. (Color online) New Fellows of the ASA. Photo by Kenny Crookston.

FIG. 7. (Color online) Scott D. Sommerfeldt, Chair of the Salt Lake City meeting. Photo by Kenny Crookton.

Events/Meeting Administrator; Kent L. Gee, Craig C. Smith, Audio-Visual; Lisa B. Sommerfeldt, Accompanying Persons Program; Scott L. Thomson, Signs/Publicity; Benjamin M. Faber, Technical Tour; Jonathan D. Blotter, Posters; Scott D. Sommerfeldt, Meeting Room Coordinator; William J. Strong, Fellows Luncheon. He also expressed thanks to the members of the Technical Program Organizing Committee: Timonty W. Leishman, Technical Program Chair; Matthew A. Dzieciuch, Acoustical Oceanography; Gary J. Rose, Animal Bioacoustics; Bruce C. Olson, Architectural Acoustics; Michael R. Bailey, Biomedical Ultrasound/ Bioresponse to Vibration; William J. Strong, Education in Acoustics and Musical Acoustics; James M. Powers, Engineering Acoustics; Connor Duke, Noise; Kent L. Gee, Richard Raspet, Physical Acoustics; Donal G. Sinex, Psychological and Physiological Acoustics; Sean K. Lehman, Karl A. Fisher, Signal Processing in Acoustics; Bruce L. Smith, Shawn Nissan, Rachel Hayes-Harb, Speech Communication; Craig C. Smith, Structural Acoustics and Vibration; Tracianne Nielsen, Underwater Acoustics.

The Plenary Session concluded with the presentation of the Vice President's gavel to Whitlow Au and the President's Tuning Fork to Anthony Atchley, in recognition of their service to the Society during the past year (see Figs. 8 and 9).

FIG. 8. (Color online) George Frisk, ASA Vice President-Elect (r) presents gavel to Whitlow Au, Vice President (l). Photo by Kenny Crookston.

FIG. 9. (Color online) President-Elect Gilles Daigle (l) presents Presidents' Tuning Fork to Anthony Atchley (r). Photo by Kenny Crookston.

The full technical program and award encomiums can be found in the printed meeting program or online for readers who wish to obtain further information about the Salt Lake City meeting (visit scitation.aip.org/jasa and select Volume 121, Issue 5, from the list of available volumes).

We hope that you will consider attending a future meeting of the Society to participate in the many interesting technical events and to meet with colleagues in both technical and social settings. Information about future meetings can be found in the *Journal* and on the ASA Home Page at ⟨http://asa.aip.org⟩.

ANTHONY A. ATCHLEY
President 2006—2007

Annual Report of the Technical Committee on Musical Acoustics

(See September and November issues for additional reports)

During 2006–2007 the Technical Committee on Musical Acoustics (TCMU) was chaired by James Beauchamp. Representatives to ASA committees were: James P. Cottingham, Membership; Uwe Hansen, Medals and Awards; Ian M. Lindevald, ASACOS; and James W. Beauchamp, Technical Council. Associate Editors were Diana Deutsch and Neville H. Fletcher. Technical Program Organization Committee (TPOM) representatives were Shigeru Yoshikawa (Honolulu) and William J. Strong (Salt Lake City). Those appointed or reappointed for 2007–2010 terms as members of TCMU are R. Dean Ayers, Judith C. Brown, Courtney B. Burroughs, John R. Buschert, Joel Gilbert, Thomas M. Huber, Bozena Kostek, Barry Larkin, Daniel O. Ludwigsen, Thomas R. Moore, Thomas D. Rossing, David B. Sharp, Julius O. Smith, William J. Strong, and Joseph Wolfe.

At the Joint Meeting with the Acoustical Society of Japan (ASJ) in November–December, 2006, TCMU in conjunction with the ASJ presented four special sessions: "Comparing Asian and Western Instruments I & II," organized and cochaired by Tomoyasu Yaguti (ASA) and Thomas D. Rossing (ASA); "Acoustical Correlates of Timbre in Music and Musical Instruments," organized and cochaired by Masashi Yamada (ASJ) and James W. Beauchamp (ASA); "Simulation and Measurement Techniques for Musical Acoustics I & II," organized and cochaired by Shigeru Yoshikawa (ASJ), Murray D. Campbell (ASA) and Gary Scavone (ASA); and "Music Information and Communication," organized and cochaired by Masuzo Yanagida (ASJ) and Bozena Kostek (ASA). In addition, TCMU was the cosponsor of the Psychological and Physiological Acoustics sponsored special session "Perception of Music and Speech: Evidence for Modularity and for Common Processing Mechanisms," which was organized and cochaired by Diana Deutsch (ASA) and Yoshitaka Nakajima (ASJ) and a cosponsor of the Architectural Acoustics sponsored session "Special Session in Honor of Manfred Schroeder," which was organized and cochaired by Ning Xiang (ASA), Juergen Schroeter (ASA), and Akira Omoto (ASJ). Moreover, there were three sessions of contributed papers at this meeting: "Music Performance," chaired by Lydia Ayers (ASA) and Akira Nichimura (ASJ); "Singing Voices," chaired by Peter Hoekje (ASA) and Naotoshi Osaka (ASJ); and "Musical and Structural Aspects of Instruments," chaired by Nicholas J. Giordano (ASA) and Isoharu Nichiguchi (ASJ). Immediately after the "Comparing Asian and Western Instruments II" session a delightful mini-concert of koto music was performed by Tamaki Ando, who proved to be an outstanding virtuoso on this instrument.

Four special sessions were presented at the Salt Lake City meeting in June, 2007: "Voice Production and Pedagogy," organized and chaired by Ingo Titze (this was followed by a mini-concert of "Yodeling as a Style of Vocal Production," presented by professional yodeler Kerry Christensen); "Flow Dynamics in Musical Acoustics," organized and chaired by Rolf Bader; "Myths and Mysteries in Musical Acoustics," organized and cochaired by Evan B. Davis and Thomas R. Moore; "Musical Requirements for Western and Non-Western Worship Spaces," organized and cochaired by Paul T. Calamia and Ewart A. (Red) Wetherill. Also, George A. Bissinger chaired "General Topics in Musical Acoustics," consisting of contributed papers. In addition, TCMU cosponsored the TCAA special session "Effects of Rooms on the Sound of Organs," organized and chaired by Bruce C. Olson (this took place at the LDS Conference Center and included a short organ concert by John Longhurst) and contributed a Hot Topics talk by Kelly Fitz entitled "Rehabilitating the world's worst sounds: Time-frequency reassignment in sound modeling and morphing." Immediately preceding the Hot Topics session William J. Strong, a long-time active TCMU member, gave the Acoustics Education Prize Lecture "Descriptive

acoustics of music and speech." The Monday evening Tutorial Lecture was given by another long-time active TCMU member, Uwe J. Hansen, who presented "Musical Acoustics: Science and Performance," a talk/demonstration that heavily depended on the performance artistry of the Salt Lake City Jazz Orchestra and drew in many Salt Lake City local students who were invited to attend.

TCMU continues to promote student involvement in musical acoustics. Since the Honolulu meeting, Eric A. Dieckman, an undergraduate student from Truman State University, has been TCMU's representative to the Student Council. Also, TCMU has continued to sponsor the ASA Best Student Paper Awards in Musical Acoustics. There were ten entries for the competition at the Honolulu meeting and two entries for the Salt Lake City meeting. The winners for Honolulu were Ricardo da Silva, who received First Place for "A hybrid approach for simulating clarinetlike systems involving the lattice Boltzmann method and a finite difference scheme," and Eric A. Dieckman, who won Second Place for "Input impedance of Asian free-reed organs." At Salt Lake City the winners were Brian B. Monson, First Place for "The 1:6 ratio in vocal pedagogy," and John Anderson Mills, III, Second Place for "Application of cochlear analysis techniques to finding percussive events in electro-acoustic music."

A number of papers on musical acoustics topics were published in JASA during 2006–2007. There were papers on such topics as the acoustics of the violin, the flute, the clarinet, the harp, the piano, the xylophone, the Australian didjeridu, and the singing voice.

JAMES BEAUCHAMP
Chair

ADVANCED-DEGREE DISSERTATION IN ACOUSTICS

Editor's Note: Abstracts of Doctoral and Master's theses will be welcomed at all times. Please note that they must be limited to 200 words, must include the appropriate PACS classification numbers, and formatted as shown below. If sent by postal mail, note that they must be double spaced. The address for obtaining a copy of the thesis is helpful. Submit abstracts to: Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747–4502, e-mail: asa@aip.org.

Cross-language study of age perception [71.Bp, 71.Hw]—Kyoko Nagao, Department of Linguistics and Department of Speech and Hearing Sciences, Indiana University, Bloomington, IN. September 2006 (Ph.D.).

This research examined the effects of listener's language familiarity on the perception of a talker's age in three linguistic contexts varying the amount of information. Two groups of listeners (English and Japanese) estimated the age of talkers whose native languages were matched or mismatched with the listener's. Furthermore, in order to investigate the effect of age stereotypes in each language, the same listeners estimated the age of talkers who disguised themselves as 20 years older or younger than their age. Results indicated that a listener's estimation of a talker's age improved when more information was available. The listeners estimated the talker age more accurately in the familiar language than the foreign language. Better age estimation was found for female talkers than male talkers, but the effect of talker's sex only appeared in the age estimation in the familiar language. Results of age estimation for age-disguised speech revealed that both language groups in this investigation have similar age stereotypes. These results suggest not only that there exists an underlying perceptual mechanism for identifying the age that is common across languages, but also suggest that the perception of age is based on language-specific variation that is accessible only to a native listener.

Advisors: Kenneth de Jong and Diane Kewley-Port.

Meetings Calendar

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

2007

22–24 Oct   NOISE-CON 2007, Reno, NV [Institute of Noise Control Engineering, INCE Business Office, 210 Marston Hall, Ames, IA 50011-2153, Tel.: 515-294-6142; Fax: 515-294-3528; E-mail: ibo@inceusa.org].

27 Nov–1 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

29 June–4 July	Acoustics08, Joint Meeting of the Acoustical Society of America (ASA), European Acoustical Association (EAA), and the Acoustical Society of France (SFA), Paris, France [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; www: http://asa.aip.org/meetings.html].
2–6 July	International Clarinet Association Clarinetfest®2008, Univ. of Missouri—Kansas City, MO [Dr. John Cipolla, Department of Music, Western Kentucky University, 1906 College Heights Blvd. #41029, Bowling Green, KY 42101-1029; Tel.: 270-745-7093; E-mail: john.cipolla@wku.edu].
27–30 Jul	NOISE-CON 2008, Dearborn, MI [Institute of Noise Control Engineering, INCE Business Office, 210 Marston Hall, Ames, IA 50011-2153, Tel.: 515-294-6142; Fax: 515-294-3528; E-mail: ibo@inceusa.org].
28 Jul–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; 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).

Revision List

New Associates

Abercrombie, Clemeth L., 1204 NW 20th Ave., #305, Portland, OR 97209.

Adcock, James L., 5005 155th Pl., SE, Bellevue, WA 98006.

Bard, Seth E., IBM, M/S P226, 2455 South Rd., Bldg. 704, Boardman Rd. Site, Poughkeepsie, NY 12601.

Beard, James K., 122 Himmelein Rd., Medford, NJ 08055.

Beard, Paul C., Medical Physics, University College London, Gower St., London WC1E 6 BT, UK.

Chang, Enson, Applied Signal Technology, Inc., Ocean Systems Div., 21311Hawthorne Blvd., Ste. 300, Torrance, CA 90503.

Cichock, Joseph A., Northrop Grumman, SD&T, 2000 West NASA Blvd., Melbourne, FL 32902.

Clark, Cathy Ann, Sensors & Sonar Systems Dept., NUWCDIVNPT, 1176 Howell St., Newport, RI 02841.

Cochran, Sandy, Univ. of Paisley, Microscale Sensors, School of Engineering and Science, High St., Paisley PA1 2BE, Scotland.

Cross, Emily L., Cavanaugh Tocci Associates, Inc., 327F Boston Post Rd., Sudbury, MA 01776.

Curra, Francesco P., Ctr. for Industrial and Medical Ultrasound, Applied Physics Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105.

Damljanovic, Vesna, Biomedical Engineering, Boston Univ., 44 Cummington St., Boston, MA 02215.

de Groot-Hedlin, Catherine D., Scripps Inst. of Oceanography, Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0225.

Derveaux, Gregoire P., INRIA, Domaine de Voluce Av., BP 105, Le Chesnay 78153, France.

Fisher, David A., 187 Highland Ave., Floor 1, Somerville, MA 02143.

Fleischman, Robert, Soning Praha, A.S., Plzenska 66, Prague 151-24, Czech Republic.

Fristrup, Kurt M., Natural Sounds Program, National Park Service, 1201 Oakrdige Dr., Ste.100, Fort Collins, CO 80525.

Golay, Francis, Lab. Regionale des Strasbourg, G5 (Acoustique), 11 rue Mentelin, Strasbourgh 67035, France.

Greenwood, William A., Pacific Lutheran Univ., 121st and Park Ave., Tacoma, WA 98447.

Han, Jin-Ha, Prosonic Co. Ltd., 71-6 Sinpeung-Ri, Geunchen-Eub, Gyongju-Si, Gyongbuk 780-900, South Korea.

Harrison, Patrick W., Sound Fighter Systems, LLC, 6135 Linwood Ave., Shreveport, LA 71106.

Hofbeck, Mary, 5622 180th St., SW, Lynnwood, WA 98037.

Huang, Lianjie, Los Alamos National Lab., MS D443, Los Alamos, NM 87545.

Jacobs, Jodi, Lencore Acoustics Corp., 1 Crossways Park Dr., West, Woodbury, NY 11797.

Jenkins, Adam C., The Greenbusch Group, 1900 West Nickerson St., Ste. 201, Seattle, WA 98119.

Kaufman, Jay R., Kaufman and Associates, 5832 Burnet Ave., Van Nuys, CA 91411.

Kautzman, Craig S., The Tennant Company, 701 North Lilac Dr., Minneapolis, MN 55422.

Ko, Wing P., Transportation Business Group, CH2M Hill, 555 South Flower St., Ste. 3550, Los Angeles, CA 90071.

Kollevoll, Kristan G., BRD Noise and Vibration Control, Inc., 112 Fairview Ave., P.O. Box 127, Wind Gap, PA 18091-0127.

Lauback, Scott, Cambridge Sound Management, 33 Moulton St., Cambridge, MA 02138.

Leonard, Jonathan S., Lencore Acoustics Corp., 1 Crossways Park Dr., West, Woodbury, NY 11797.

Li, Zhaung, 900 Pump Rd., Apt. 76, Richmond, VA 23238.

Luo, Xin, Dept. of Auditory Implants and Perception, House Ear Inst., 2100 West Third St., Los Angeles, CA 90057.

Mukdadi, Sam M., Mechanical and Aerospace Eng., West Virginia Univ., P.O. Box 6106, Morgantown, WV 26506-6106.

Nusbaum, Howard C., Dept. of Psychology, Univ. of Chicago, 5848 South University Ave., Chicago, IL 60637.

Otani, Makoto, Dept. Intelligent Systems Design Eng., Toyama Prefectural Univ., Kurokawa 5180 Imizo, Toyama 939-0398, Japan.

Qin, Shengping, Biomedical Engineering, Univ. of California, Davis, One Shields Ave., Davis, CA 95616.

Saucier, Scott G., Tibbetts Industries, Inc., Colcord Ave., P.O. Box 1096, Camden, ME 04843.

Smith, David R. R., Physiology, Development & Neuroscience, Univ. of Cambridge, Downing St., Cambridge CB2 3EG, UK.

Southall, Brandon L., US Dept. of Commerce, Natl. Oceanic and Atmospheric Admin., Natl. Marine Fisheries Service, 1315 East-West Highway, SSMC III #12539, Silver Spring, MD 20910.

Tardelli, John D., ARCON Corp., Digital Speech Processing, 150K New Boston St., Wobun, MA 01801.

Tyler, Michael D., MARCS Auditory Laboratories, Univ. of Western Sydney, Bldg. 5, Bankstown Campus, Locked Bag 1797, Penrtih South NSW 1797, Australia.

Vondrasek, Martin, Soning Praha, A. S., Plzenska 66, Prague 151-24, Czech Republic.

Walters, David T., 1344 South 7th St., Lincoln, NE 68502.

Xiao, Jianqiang, Hunter College, Psychology, 695 Park Ave., New York, NY 10021.

Yasutaka, Ueda, Hazama Corp., TRI, 515-1 Karima, Tsukuba, Ibaraki, 305-0822, Japan.

Zarnetske, Michael R., 115 Bay State Dr., Braintree, MA 02184.

New Students

Abada, Shani H., 3601 Sainte Famille, Apt. 803, Montreal QC H2X 2L6, Canada.

Anguah, Kofi A., Swarthmore College, 500 College Ave., Swarthmore, PA 19081.

Arz, Jean-Pierre, 5011 la Fontaine, Montreal QC H1V 1R6, Canada.

Azzara, Alyson J., 401 Anderson St., #8A, College Station, TX 77840.

Babel, Molly E., Linguistics, Univ. of California, Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720-2650.

Barroso, Celia, 3556 Maplewood Ave., Los Angeles, CA 90066.

Baumgart, Johannes, Inst. for Aerospace Eng., Technische Univ. Dresden, Dresden 01062, Germany.

Beckmann, Daniel F., 5507 Rosewood St., Roeland Park, KS 66205.

Benson, David H., 1610 Sherbrooke St., W., Apt. 112, Montreal QC H3H 1E1, Canada.

Berkowitz, Michael J., Psychology Dept., Vanderbilt Univ., 301 Wilson Hall, 111 21st Ave., South, Nashville, TN 37203.

Blumenrath, Sandra H., Dept. of Biology and Psychology, Univ. of Maryland, College Park, MD 20742.

Bodson, Anais, Universit Bochum, Universitatsstrasse ND 6/33, Bochum D-44780, Germany.

Bradonikic, Kaca, 311 Allston St., Apt. 12, Brighton, MA 02135.

Camacho, Arturo, 700 SW 16th Ave., #108, Gainesville, FL 32601.

Chang, Chiung-Yun, Speech and Hearing Science, Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210.

Cheng, Rui, Aerospace Engineering, Penn State Univ., 225A Hammond Bldg., University Park, PA 16802.

Choi, James J., 379 Windsor Rd., Englewood, NJ 07631.

Cholewiak, Danielle M., Cornell Univ., Bioacoustics Research Program, 159 Sapsucker Woods Rd., Ithaca, NY 14850.

Clark, Mehgan, 3-212 Lamoine Village, Macomb, IL 61455.

Collin, Jamie R., Magdalen College, High St., Oxford OX1 4AU, UK.

DeAngelis, Chris, Box 20584, Cranston, RI 02920.

Derezinsk, Steve J., 1008 Mass Ave., Apt. 702, Cambridge, MA 02138.

DiZinno, Nicholas, 1094 Reilly St., Bay Shore, NY 11706.

Dossot, Georges A., Ocean Engineering, Univ. of Rhode Island, Sheets Bldg., Narragansett Bay Campus, Narragansett, RI 02882.

Duke, Jessica E., Psychology Dept., Emory Univ, 532 Kilgo Cir., Atlanta, GA 30322.

Feizollahi, Zhaleh, Linguistics, Georgetown Univ., 37th & "0" St., Washington, D.C. 20057.

Foale, Cameron B., 808 Tress St., Ballarat VIC 3350, Australia.

Gaston, Jeremy R., Psychology Dept., Binghamton Univ., Binghamton, NY 13901.

Gauthier, Bruno, 2831 Jenne d'Arc, Apt. 202, Montreal QC H1W 3V8, Canada.

Giannos, Evangelia, 217 South St., Jamaica Plain, MA 01230.

Gomes, Maria L., FACINTER Faculdade Internacional, Avenida Luiz Xavier 103, Curitiba, Parana, 80021-980 Brazil.

Gregg, Mellisa K., 333 Candee Ave., Apt. G4, Sayville, NY 11782.

Harvey, Ryan B., BAE Systems, 1250 24th St., NW, Ste. 850, Washington, D.C. 20037.

Hearst, Jason, 30 St. Stephen St., Boston, MA 02115.

Hon, Elisabeth, MIT, 60 Wadsworth St., Apt. 19E, Cambridge, MA 02142.

Hsieh, I-Hui, 5513 Verano Pl., Irvine, CA 92617.

Joshi, Aditya, 17C Bayberry Rd., New Bedford, MA 02740.

Kenny, Ryan J., Psychology Dept., Boston College, 301 McGuinn Hall, 140 Commonwealth Ave., Chestnut Hill, MA 02467.

Kracht, Jonathan M., 295 East Church St., Sellersville, PA 18960.

Kumpulanian, Danielle, 7 Country Ln., Palmer, MA 01069.

Lai, Puxiang, Aerospace and Mechanical Eng., Boston Univ., 110 Cummington St., Boston, MA 02215.

Law, II, Franzo F., 1337 Jefferson Ave., Brooklyn, NY 11221.

Lee, Ji-Yeoun, R405, 103-6, Munji-dong, Yuseong-gu, Daejeon 305-732, Republic of Korea.

Leger, Mark L., 8 Charles Plaza, Apt. 209, Baltimore, MD 21201.

Lezamiz, Lucas, Applied Physics Lab., Univ. of Washington, 1013 NE 40th St., Box 355640, Seattle, WA 98105-6698.

Li, Naihsin, National Taiwan Univ., Graduate Inst. of Linguistics, No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan.

Lin, Wan, 337795 Gatech Station, Atlanta, GA 30332.

Lin, Yao-ju, English Dept., Linguistic Div., National Taiwan Normal Univ., 162 HePing East Rd., Sec. 1, Taipei 106, Taiwan.

Liu, Fei, 282 Corry Village, Apt. 7, Gainesville, FL 32603.

Liu, Sheng, 865 Tucker Rd., Apt. 4, North Dartmouth, MA 02747.

Lopez, Karece, 90-79 180th St., Jamaica, NY 11432.

Love, Katharine L., 4015 Hudson Dr., Hoffman Estates, IL 60195.

Luo, Haibaio, Aerospace and Mechanical Eng., Boston Univ., 110 Cummington St., Boston, MA 02134.

MacLeod, Mark H., Mechanical Engineering, Johns Hopkins Univ., 3400 North Charles St., Batimore, MD 21218.

Maleke, Caroline, Biomedical Engineering, Columbia Univ., New York, NY 10027.

Mancini, Jolene A., Hearing, Speech, and Language Science, Gallaudet Univ., 800 Florida Ave., NE, Washington, D.C. 20002.

Maruska, Karen P., Hawaii Inst. of Marine Biology, 46-007 Lilipuna Rd., Kaneohe, HI 96744.

McCarty, Candice Q., 3610 Driftwood Dr., Lafayette, IN 47905.

Meyer, Matthias, Gartenstrasse 13, Wiefelstede 26215, Germany.

Mihalcik, Ladislav, National Healthy Organization, NRC Noise and Vibration, RUZ Bratislava, Ruzinovska 8, Bratislava, Slovakia 829 09, Slovak Republik.

Miller, Denise M., 1013 Old Boalsburg Rd., Apt. 6, State College, PA 16801.

Moallem, Theodore M., RLE Sensory Communication Group, Massachusetts Inst. of Tech., 77 Massachusetts Ave., Rm. 36-737, Cambridge, MA 02139.

Navaladi, Akshay, Biomedical Engineering, Boston Univ., 44 Cummington St., Boston, MA 02215.

Newman, Kelly A., School of Fisheries and Ocean Sciences, Univ. of Alaska Fairbanks, 245 Oneill Bldg., P. O. Box 757220, Fairbanks, AK 99775-7220.

Noirot, Isabelle C., Charles Plaza 8, North Tower, Apt. 1303, Baltimore, MD 21201.

Noisternig, Markus, Inst. of Electronic Music & Acoustics, Univ. of Music and Dramatic Arts, Inffeldgasse 10/3, Graz 8010, Austria.

Olmstead, Anne J., 69 Varga Rd., #112, Ashford, CT 06278.

Oorellana, Douglas W., 104 West Univ. Parkway, Apt. B1, Baltimore, MD 21210.

Pogal-Sussman, Tracy, BME Dept., Boston Univ., 44 Cummington St., Boston, MA 02215.

Pruthi, Tarun, 9314 Cherry Hill Rd., 917, College Park, MD 20740.

Quijano, Jorge E., 2139 West Burnside St., Apt. 203, Portland, OR 97210.

Roberts, Paul L. D., Electrical and Computer Engineering, Univ. of California, San Diego, 9500.

Gilman Dr., La Jolla, CA 92093-0238.

Rosenbaum, Joyce E., 2 Horizon Rd., PH1, Fort Lee, NJ 07024.

Rossi-Katz, Jessica, 112 Kolar Court, Erie, CO 80516.

Rousounelos, Andreas, Flayt 147, Matthias Court, Silk St., Salford M3 6JF, UK.

Rowland, Sarah A., Psychology Dept., Univ. of Connecticut, 406 Babbidge Rd., Unit 1020, Storrs, CT 06269-1020.

Sadaka, Janine, 9 Jeffrey Ln., Great Neck, NY 11020.

Sathyendra, Harsha M., Electrical Engineering, Univ. of Florida, 216 Larsen Hall, Gainesville, FL 32611.

Savitala, Hari V., 405 3rd St., #2, Troy, NY 12180.

Schneider, Jennifer N., Psychology Dept., SUNY Buffalo, Park Hall, Room 206, Buffalo, NY 14260-4110.

Son, Minjung, Haskins Labs., 300 George St., #900, New Haven, CT 06511.

Steen, Thomas L., Mechanical Eng., Boston Univ., 110 Cummington St., Boston, MA 02215.

Threw, Barry J., 2215 Ward St., Berkeley, CA 94705.

Tibrea, Roxana D., 20 Fairway Dr., East Hampton, NY 11937.

Trout, Justin N., 196-1 Allston St., Allston, MA 02134.

Uysal, Ismail, Electrical and Computer Eng., Univ. of Florida, 216 Larsen Hall, Center Dr., Gainesville, FL 32611.

Van Ark, Emily M., MIT-WHOI, Geology and Geophysics, 77 Massachusetts Ave., 54-517A, Cambridge, MA 02139.

Vasques, Cesar M. A., DEMEGI/FEUP, Rua Dr. Roberto Frias, s/n Edificio M, Sala 206, Porto 4200-465, Portugal.

Viswanathan, Navin, 90a Birch St., Willimantic, CT 06226.

Wang, Weixiong, Dept. of Engineering Mechanics, Fluid Acoustics Lab., Tsinghua Univ., Beijing 100084, China.

Wiessner, Nicole L., 2010 Misty Hollow Court, Forney, TX 75126.

Wilkason, Colby, 120 Fairways Dr., Warner Robins, GA 31088.

Yarbrough, Ray A., Applied Research Labs., Univ. of Texas, 10000 Burnet Rd., Austin, TX 78758.

New Electronic Associates

Aura, Matias, Wartsila Finland Oy, Calculations and Simulations, Jarvikatu 2-4, Vaasa FI-65101, Finland.

Boehme, Hollis, 320 King Arthur Court, Austin, TX 78746.

Ess, Robert H., 1107 Pine Hollow Dr., Friendswood, TX 77546.

Garrido Lopez, David, Musicos 26, Tres Cantos, Madrid 28760, Spain.

Gazagnaire, Julia, Naval Surface Warfare Ctr., Dept. of Defence, 110 Vernon Ave., Panama City Beach, FL 32407.

Greenlee, Doug, 995 South Garfield St., Denver, CO 80209.

Griffin, Sararh J., MRC Toxicology Unit, Univ. of Leicester, P.O. Box 138, Hodgekin Bldg., Lancaster Rd., Leicester LE1 9HN, UK.

Hain, Tim C., Northwestern Univ., 645 N. Michigan, Chicago, IL 60611.

Heintze, Olaf, Inst. of Composite Structures and Adaptive Systems, German Aerospace Ctr., Lilienthalpaltz 7, Braunschweig 38108, Germany.

Hetzer, Claus, National Ctr. for Physical Acoustics, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677.

Huang, Caroline B., 39 Howells Rd., Belmont, MA 02478.

Lehman, Mark E., Communication Disorders, Central Michigan Univ., 2173 Health Professional Bldg., Mt. Pleasant, MI 48859.

Maetani, Toshiki, Otolaryngology/Head and Neck Surgery, Ehime Univ. School of Medicine, Shitsukawa, Toon, Ehime 791-0295, Japan.

Mahajan, Sanjay K., 28834 W. King William Dr., Farmington Hills, MI 48331.

Mansky, James M., Earth Tech, One World Financial Ctr., 200 Liberty St., 25th Floor, New York, NY 10281.

Marquez, Ronald, IDS, 353 James Record Rd., Huntsville, AL 35801.

Nielsen, Johan L., Tandberg, Philip Redersens vei 22, Lysaker 1355, Norway.

Ono, Nobutaka, Dept. Information Physics and Computing, Grad. Sch. of Info. Sci. and Tech., The Univ. of Tokyo, 7-3-1 Hongo, Tokyo 113-8656, Japan.

Pastuszek-Lipinska, Barbar E., ul. Iglasta 8, Grotniki-Ustronie, Poland 95-073.

Pfaffli, Gregor Grischa, Gerberngasse 21A, Bern 3011, Switzerland.

Quinn, Sandra, Psychology Dept., Univ. of Stirling, Stirling FK9 4LA Scotland, UK.

Ramprashad, Sean A., DoCoMo USA Labs, Media Lab., 181 Metro Dr., San Jose, CA 95110.

Serkhane, Jihene E., Cold Spring Harbor Lab., Freeman Bldg., 1 Bungtown Rd., Cold Spring Harbor, NY 11724.

Shah, Gaurav S., Eliza Corp., 100 Cummings Ctr., Ste. 348G, Beverly, MA 01915.

Slaughter, Julie C., Etrema Products, Inc., 2500 North Loop Dr., Ames, IA 50010.

Tregenza, Nicholas J., 5 Beach Ter., Long Rock, Cornwall, TR20 8JE UK.

Vitchev, Nikolay V., 4363 Cherry Ave., San Jose, CA 95118.

Whittum, David H., Varian Medical Systems, Microwave and Physics R&E, 911 Hansen Way, C077, Palo Alto, CA 94304.

Wilson, M. Lee, Shimoda Environmental, Inc., 7602 Stoneywood Dr., Austin, TX 78731.

Yokotani, Yoshikazu, Westliche Stadtmauerstr. 38, Zimmer #9, Erlangen, Bayern, 91054, Germany.

New Corresponding Electronic Associates

Bistafa, Sylvio R., Dept. of Mechanical Eng., Univ. of Sao Paulo, Av. Prof. Melo Moraes 2231, Sao Paulo SP 05508-900, Brazil.

Daunys, Gintautas, Siauliai UIniv., Electronics, Vilniaus 141, Siauliai 76353, Lithuania.

Meesawat, Kittiphong, Dept. of Electrical Eng., Khon Kaen Univ., 123 Mitraphab Rd., Mueng, Khon Kaen, 40002, Thailand.

Pal, Amita, Bayesian & Interdisciplinary Research Unit., Indian Statistical Inst., 203 Barrackport Trunk Rd., Kolkata, West Bengal, 700108, India.

Srinivasan, K., Apt. 009, Admiralty Manor 83, 6th Main, Indira Nagar II Stage, Bangalore, Karnataka, 560008, India.

Szigetvari, Andrea, Berzsenyi u. 14/A, Dunakeszi, Pest, 2120, Hungary.

Members Elected Fellows

P. Blanc-Benon, D. A. Conant, A. W. Gummer, C. W. Holland, J. E. Kreiman, K. D. LePage, J. A. McAteer, D. R. Palmer, M. G. Prasad, P. A. Rona, M. Vorlaender, J. Vos

Associates Elected Fellows

S. Beristain, A. C. Gade, H. Riquimaroux, M. V. Trevorrow, B. T. Zin

Associates Elected Members

J. Ahlstrom, S. B. Blaeser, C. E. Hughes, G. E. Jacobs, X. Jiang, R. E. Kumon, B. R. Munson, S. M. Perron, T. S. Talayman

Students to Associates

W. C. K. Alberts, II, R. B. Astrom, J. G. Bernstein, L. K. Bornstein, J. L. Hiatt, A. Roginska

Students to Electronic Associates

M. S. Allen, L. B. Berry, P. A. Cariani, S. A. Cheyne, S. F. Disner, B. L. Engdahl, M. J. Epstein, W. T. Fitch, C-F. Huang, A. J. Morgan, I. Paek, A. J. Subkey, S. J. van Wijingaarden

Resigned

J. R. Ison, M. R. Noble—*Members*

R. Eklund—*Associate*

C. Lin—*Student*

Y. Igarashi—*Electronic Associate*

Deceased

S. Buus, L. Lisker, T. Litovitz, W. A. Watkins—*Fellows*

J. G. Harris, B. H. Pasewark—*Members*

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

1851

Dropped

D. Berman, A. L. Dancer, R. M. Fitzgerald, H. Himelblau, T. R. Horrall, D. S. Hougland, R. D. Kent, R. J. Porter, E. W. Sapp, E. Shang, R. Wei—*Fellows*

R. Aubauer, D. E. Ballesty, N. G. Bibikov, J. E. Bollinger, M. W. Bonilha, L. Brancazio, R. L. Brill, R. J. Brothers, V. Bucur, M. W. Burton, J. W. Caruthers, P. H. Ceperley, J. T. Christoff, T. Connelly, C. N. Corrado, T. H. Dat, T. Deierlein, F. Farhang, T. A. Frank, E. L. Goshorn, L. C. Gray, J. Guo, R. Hill, R. L. Jennette, V. B. Johnson, R. K. Kataoka, C. A. Krajewski, R. Kumaresan, C. T. Kung, J. T. Kunio, S. Kuwahara, S. Lakatos, R. A. Lee, T. Lepoint, F. Lin, J. C. Lucero, L. Luker, R. D. Magnuson, P. A. Manning, D. A. McCurdy, M. Mosher, J. G. Mottley, J. E. Murphy, M. K. Myers, H. Nakashima, M. J. Oslac, S. J. Page, J. Pan, S. Peng, P. Pernod, J. L. Postlewait, J. P. Powers, V. L. Preobrazhensky, W. A. Pritchard, M. L. Readhead, C. Reichmuth, D. C. Ricks, T. H. Rose, M. N. Rossi, D. W. Saenz, X. Serra, T. Sola, J. L. Spiesberger, Y. M. Szymko-Bennett, A. M. Terry, J. A. Thomas, K. A. Traub, H. A. Vosbein, G. Wade, G. G. Wan, R. Wilkes, D. W. Winfield, S. A. Wood, S. Wu, H. Yasuda, D. J. Zajac—*Members*

H. Arndt, L. L. Auther, M. D. Azevedo, P. Ball, M. W. Banks, A. L. Barco, D. F. Barker, T. J. Barnhart, M. C. Barron, S. M. Bellinger, J. R. Benki, L. J. Berdish, O. S. Blacklock, T. T. Block, W. H. Boober, G. R. Boxhorn, J. T. Brooks, E. D. Brown, G. C. Burnett, T. Chan, L. Cheng, S. K. Chi, P. M. Chute, F. S. Chwieroth, S. Cingolani, P. A. Clements, C. G. Clinard, W. F. Conklin, D. C. Copley, M. P. Coughlin, J. Crouch, J. A. Cynx, M. J. Daley, M. S. Datum, K. Davis, D. Donlagic, B. M. Dunkin, S. H. Eaton, R. A. Edry, H. S. Erskine, G. Evangelista, H. H. Farag, R. F. Fehlmann, J. J. Flaherty, A. A. Freeman, K. L. Frost, M. Fruhmann, D. J. Gaddis, M. R. Gahler, E. A. Gardner, E. M. Giddens, R. G. Gilmore, K. Gipson, L. R. Goldberg, S. Gravel, R. J. Greaves, M. G. Greene, P. Gudmundson, S. A. Harner, L. A. Hart, H. G. Hermann, W. J. Hess, C. J. Hickey, J. W. Hillenbrand, K. K. Ho, T. L. Hoffmann, W. Hongbin, W. N. House, C. Huang, D. H. Hurley, D. Hustvedt, S. Imai, S. D. Isard, A. Jahn, J. Jeon, G. Jinhoi, D. L. Jones, Y. Kang, C. U. Karlsson, J. B. Kelly, G. Kim, K. Kim, I. Kim, J. E. King, H. B. King, R. J. Koeppel, H. Lai, S. M. Lashomb, P. Lee, R. Lyons, M. C. Mah, S. Marburg, M. M. Matlo, R. Mazza, A. M. McCall, M. S. McPherson, C. A. Mead, T. Men, R. J. Mihelich, J. Miralles, J. A. Miranda, M. A. Mora, M. L. Moran, M. T. Mosheh, K. Nakayama, M. Navarri, M. L. Norris, M. J. O'Donnell, B. O'Sickey, B. O'Toole, S. Oh, K. Ohta, M. P. Olivieri, M. P. Orlowski, N. C. Otto, I. L. Panzer, C. A. Peak, C. V. Pecorari, A. L. Pittman, K. M. Powell, R. Preston, M. M. Prince, J. C. Pruitt, J. A. Quinlan, G. H. Recanzone, J. A. Rice, A. K. Rogers, W. Rolshofen, M. M. Rose, S. Roth, J. Roy, K. Samo, A. Savage, K. Sawicki, H. Schissel, W. J. Schultz, R. D. Simons, S. A. Singh, K. A. Siren, E. Skowbo, B. E. Smith, B. L. Southall, J. K. Spiewla, G. R. Steffen, R. L. Storms, A. D. Sweeney, D. M. Sykes, H. Takaso, D. T. Talkin, S. L. Tangonan, Q. Thompson, R. Todd, J. V. Troyan, F. M. Vale, W. M. Vanhuffelen, M. Wakumoto, C. P. Walls, M. A. Walton, Y. Wang, G. Watkins, S. A. Weil, A. T. Wells, L. A. Wilen, S. E. Wiley, T. A. Wilson, E. Wolfe, G. Wood, W. E. Worman, C. Wright, X. Wu, M. Zahui, Y. Zhang, Y. Zheng, W. R. Zimmerman—*Associates*

T. L. Arbogast, B. R. Arslan, R. Balachandran, I. Bekker, S. F. Bird, A. Birsan, R. A. Buchanan, D. Chelidze, Y. Chen, J. J. Christensen, R. F. Coulston, J. R. Davies, D. S. Davison, K. J. Delaney, B. J. Doty, A. B. Druzhinin, J. Dunne, R. M. Ead, T. L. Eadie, M. Z Fuka, T. Furukawa, J. M. Gamba, R. J. Gent, H. D. Goodman, K. K. Govindarajan, T. C. Grobaski, G. J. Hollich, A. J. Hundley, O. I. Ilkorur, A. Jacobs, J. H. Jeong, M. A. Jones, V. I. Jonsdottir, J. Keane, T. A. Khan, K. Kim, K. Kristiansen, M. A. Latimer, T. Lawu, L. J. Lee, S. V. Levi, P. Marvit, I. Masahiro, J. McCloud, H. Miyazaki, D. F. Moore, A. P. Moscoso, S. L. Nieukirk, L. Nieva, K. T. Nihei, B. C. Noon, C. D. Pierce, S. Plain, T. J. Potter, T. Qu, B. Reddy, A. N. Riddle, M. J. Rogan, K. Saleh, J. A. Scales, D. A. Schwartz, K. N. Shahin, T. I. Shevchenko, S. Spiegelberg, X. Sun, W. B. Swain, A. A. Terrinoni, N. F. Thomas, A. C. Toomey, U. Ucar, J. F. Upegui Fernandez, T. W. Wade, N. Weiland, W. Wilson, M. Xu, D. A. Yates, L. A. Zaleski, X. Zhang, C. F. Zipperle—*Electronic Associates*

B. A. Adam, O. M. Al-Bataineh, J. Bahng, B. J. Baker, B. N. Baker, S. L. Baldwin, M. Bedard, T. Beierholm, A. J. Belanger, S. A. Bentil, B. Bolat, P. J. Bourgeois, P. J. Brasovan, A. Breznik, S. A. Burgess, S. P. Carr, J. Chatterley, M. E. Chavez-Peon, Y. A. Chen, S. R. Colletti, S. M. Corbley, M. E. Cruikshank, B. Defreville, V. Dellwo, M. E. Denis, D. J. Derrick, C. Dillon, A. J. Doller, N. J. Donabed, K. Dorffner, A. Dosemeciler, J. Drummond, Y. Du, D. Eckert, B. G. Evans, E. L. Foley, A. R. Foxwell, B. Frederic, M. C. Freyaldenhoven, C. N. Galeano, B. J. German, B. L. Giordano, M. M. Glasbrenner, M. Gordon, M. R. Gramann, T. Gridley, E. Gros, M. Grube, R. Gutierrez, H. Ha, J. Haire, K. Hamaoui, S. Hansen, J. F. Hay, K. C. Hegewisch, A. H. Heidari, R. J. Hill, M. S. Hilton, A. A. Hood, S. M. House McKenna, A. K. Imbrie, J. K. Ingle, T. M. Jones, A. Kan, A. M. Kang, A. Kawakami, T. Kawenski, E. F. Kazemzadeh, R. F. Kesselman, K. J. Kilanski, M. Kim, B. Kim, T. J. Klasen, I. S. Kodavaty, K. K. Konde, M. R. Kornatz, E. M. Ku, B. Kwon, S. J. Lalk, M. D. Lammers, J. Lee, Y. Lee, C. Lewiston, L. Liaw, Y. K. Lin, P. A. Lipscomb, J. C. Liu, D. Liu, G. Lu, J. M. Luis, W. Luo, V. Madani, X. Mao, B. A. Martin, D. Mathias, J. Matsubara, A. D. Maxwell, S. L. McCoy, C. T. McMillan, T. Meng, P. Michal, N. Mirante, E. M. Mollman, M. T. Molnar, F. Monten, C. L. Moran, H. H. Namarvar, N. Namdaran, J. R. Nesbitt, L. S. Notestine, D. E. O'Gorman, L. C. Oliveira, S. A. Olsen, A. Ong, J. A. Ortiz, A. J. Ortmann, T. Owens, N. Pai, V. C. Palan, J. Parchment, E. Park, M. A. Peres, J. A. Pillai, J. Pinsonnault, A. Pinto, D. M. Playfair, B. B. Poon, A. F. Prieto, W. M. Quarles, J. M. Raisamo, K. L. Rasmussen, L. B. Reynolds, C. Richey, B. L. Robinette, K. M. Rosen, C. J. Rozell, M. D. Ruvalcaba, E. Sandrich, J. M. Schaefer, Y. Schlezinger, M. Schotten, S. V. Settle, P. J. Shapiro, J. Sheaffer, L. Shi, C. Shiung, W. N. Sloan, T. W. Spohnholtz, K. A. Sprague, C. S. Sullivan, D. R. Surendran, A. D. Tapia, S. V. Teles, S. D. Terry, M. E. Thompson, J. J. Thompson, R. I. Thomson, T. Thornhart, C. Tillery, A. A. Truett, D. P. Valente, E. Valmont, N. L. Vause, L. Venkatesh, F. Vogt, S. von Lueders, K. G. Walker, A. H. Walker, R. Wallace, T. Wang, C. J. Ward, J. E. Waxman, E. M. Wezensky, S. Wickstrom, E. T. Wienke, M. A. Wilcox, C. Wilks, M. J. Wittbrodt, W. P. Wong, B. J. Wygonik, G. Yun, Y. Zhang, X. Zheng, W. Zhu—*Students*

Fellows	902
Members	2173
Associates	2646
Students	1068
Electronic Associates	685
Total	7474

ACOUSTICAL NEWS—INTERNATIONAL

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

International Meetings Calendar

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

October 2007

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

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

17–18   **Institute of Acoustics Autumn Conference 2007**, Oxford, UK (Web: www.ioa.org.uk/viewupcoming.asp).

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

November 2007

14–16   **14th Mexican International Congress on Acoustics**, Leon, Guanajuato, Mexico (Fax: +52 55 5523 4742; e-mail: sberista@hotmail.com).

29–30   **Reproduced Sound 23**, The Sage, Gateshead, UK (Web: www.ioa.org.uk/viewupcoming.asp).

December 2007

6–9   **International Symposium on Sonochemistry and Sonoprocessing (ISSS2007)**, Kyoto, Japan (Web: www.j-sonochem.org/ISSS2007).

April 2008

8–11   *Oceans'08**, Kobe, Japan (Web: www.oceans08mtsieeekobe-technoocean08.org/index.cfm).

17–18   *Spring Meeting of the Swiss Acoustical Society**, Bellinzona (Tessin), Switzerland (Web: www.sga-ssa.ch).

June 2008

30–4   **Acoustics'08 Paris**: **155th ASA Meeting +5th Forum Acusticum** (**EAA**)+**9th Congrès Français d'Acoustique** (**SFA**), Paris, France (Web: www.acoustics08-paris.org).

July 2008

7–10   **18th International Symposium on Nonlinear Acoustics (ISNA18)**, Stockholm, Sweden (Web: www.congrex.com/18th_isna).

27–31   *10th Mechanics of Hearing Workshop**, Keele University, UK (Web: www.mechanicsofhearing.com).

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 2008

25–29   **10th International Conference on Music Perception and Cognition (ICMPC 10)**, Sapporo, Japan (Web: icmpcl0.typepad.jp).

September 2008

8–12   *International Symposium on Underwater Reverberation and Clutter**, Lerici, Italy (Web: isurc2008.org).

15–17   *International Conference on Noise and Vibration Engineering (ISMA2008)**, Leuven, Belgium (Web: www.isma-isaac.be).

16–18   *Underwater Noise Measurement**, Southampton, UK (Web: www.ioa.org.uk/viewupcoming.asp).

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

October 2008

21–24   **Acústica 2008**, Coimbra, Portugal (Web: www.spacustica.pt).

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

November 2008

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

September 2009

6–10   **InterSpeech 2009**, Brighton, UK (Web: www.interspeech2009.org).

August 2010

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

September 2010

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

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*
DIMITRI DONSKOY, *Stevens Institute of Technology, Castle Point on the Hudson, Hoboken, New Jersey 07030*
GEOFFREY EDELMANN, *Naval Research Laboratory Code 7145, 4555 Overlook Ave. SW, Washington, DC 20375*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627*

7,215,598

43.30.Gv IMAGING SONAR AND DETECTION SYSTEM USING SUCH A SONAR

Pierre Guthmann, assignor to Thales
8 May 2007 (Class 367/88); filed in France 13 July 2001

Yet another company has rediscovered basic beamforming and attempted to patent it. This time it involves a 2-D array that transmits and receives on different axes.—GFE

7,206,257

43.35.Ei ACOUSTIC REMOTE CAVIATION AS A DESTRUCTION DEVICE

James C. S. Meng, assignor to The United States of America represented by the Secretary of the Navy
17 April 2007 (Class 367/137); filed 2 September 2003

The author asserts that acoustic cavitation can be used as an environmentally friendly (yet extremely loud) method to destroy mines and even torpedos. Perhaps the scrubbing power of bubbles has been overestimated.—GFE

7,215,599

43.35.Yb ULTRASONIC TRANSMITTER, ULTRASONIC TRANSCEIVER AND SONAR APPARATUS

Yasushi Nishimori *et al.*, assignors to Furuno Electric Company, Limited
8 May 2007 (Class 367/138); filed in Japan 27 November 2002

This patent relates to the specifics of producing ultrasonic waveforms via a signal generator. The authors argue that ringing due to improper driving pulse signals can lead to false target detection.—GFE

7,204,146

43.35.Zc DEVICE AND METHOD FOR MEASURING THICKNESS

Ichiro Ishimaru and Takahiro Okuda, assignors to Techno Network Shikoku Company, Limited
17 April 2007 (Class 73/579); filed in Japan 5 September 2003

The patent describes a well known resonance method for measuring thickness of objects, emphasizing thickness measurements of a thin film formed on a surface of an object, such as a film of paint on an automobile body. The patent claims the use of multiple vibration detectors **13**, vibration source **10**, and signal analyzer **16**. The vibration source and detectors are
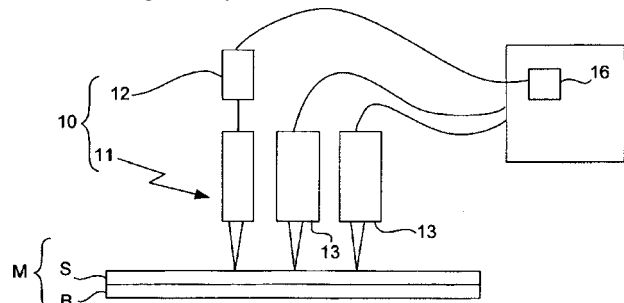
preferably noncontact optical (laser)-based. The signal analyzer determines the film's thickness resonant frequencies and speed of sound and calculates the thickness. The patent asserts the utility of the method for measuring the thickness of an individual layer in a multilayer coating.—DMD

7,210,352

43.38.Bs MEMS TEETER-TOTTER APPARATUS WITH CURVED BEAM AND METHOD OF MANUFACTURE

John S. Foster and Paul J. Rubel, assignors to Innovative Micro Technology
1 May 2007 (Class 73/514.32); filed 14 June 2005

This patent describes a new twist on the single-axis accelerometer. A beam **430** is etched out of the substrate material—and is curved upwards by evaporation of a stress-inducing film **475**. This beam is hinged on support **450** and rests over a capacitor on each end (**480a** and **480b**). A proof mass
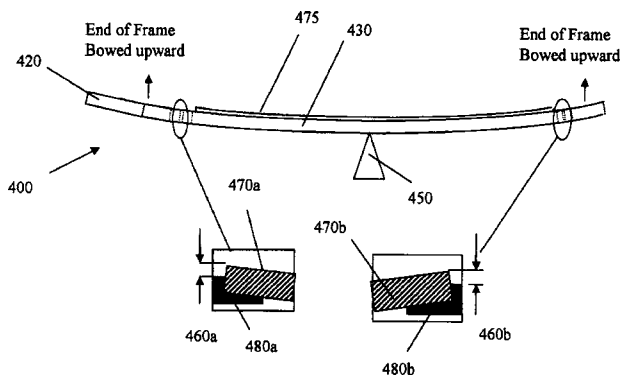
420 is added to one end to make it acceleration sensitive, unbalancing a capacitance bridge as it tips. But wouldn't such an accelerometer also be highly sensitive to rotation?—JAH

7,215,527

43.38.Bs MEMS DIGITAL-TO-ACOUSTIC TRANSDUCER WITH ERROR CANCELLATION

John J. Neumann, Jr. and Kaigham J. Gabriel, assignors to Carnegie Mellon University
8 May 2007 (Class 361/233); filed 20 September 2004

This patent discloses an electrostatic transducer that functions as both a digital loudspeaker and as a microphone. This is specifically a MEMS design incorporating a serpentine silicon spring suspension beneath a 2 ×2 mm polymer diaphragm that is applied over it. Although the patent description is quite detailed about the modeling and process steps involved, it is not clear to this reviewer that such a device has ever been built and made to work. The modeling predicts a frequency response as a loudspeaker extending to 7 kHz and a sensitivity of 84 dB for a 5 V input. There is really nothing new here besides the suspension design, and the only reason it can be called digital is because there may be a digital-to-analog converter on the same chip.—JAH

7,207,222

43.38.Fx ELECTROMAGNETIC PIEZOELECTRIC ACOUSTIC SENSOR

Michael Thompson and Scott Ballantyne, assignors to Sensorchem International Corporation
24 April 2007 (Class 73/590); filed in Canada 28 August 2001

A so-called "electromagnetic piezoelectric acoustic sensor" (EMPAS) is described in this patent. This is a piece of quartz with center-dot conducting electrodes. In the referenced application, the quartz crystal is one wall of a flow-through liquid cell into which various chemical reagents can be admitted and the resulting resonant frequency shifts of the EMPAS sensor can be measured. The quartz appears to be used in thickness shear mode operating loosely coupled to the electric field of a resonant coil. The arrangement of coil and sensor crystal seems reasonable, though one suspects that better magneto-electric coupling is possible with other coil configurations.—JAH

7,216,542

43.38.Fx THREE-COMPONENT PIEZOELECTRIC VIBRATION ACCELEROMETER PROVIDED WITH A SENSOR

Igor Borisovich Kobyakov, Moscow, Russian Federation
15 May 2007 (Class 73/514.34); filed in Russian Federation 29 November 2002

This patent discloses the existence of an optimum construction of a triaxial accelerometer using piezoelectric elements of the PZT and Zn-O
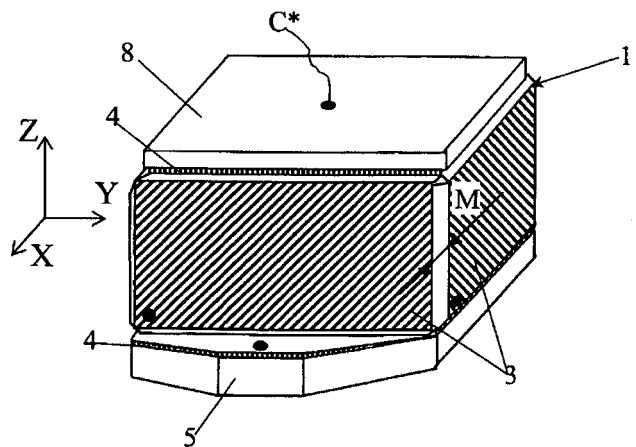
type—i.e., polycrystalline. The design utilizes combinations of all three directions of electric held to obtain triaxial sensitivity. The new information disclosed here seems to be the fact that there are optimal locations for the lead attachments.—JAH

7,177,438

43.38.Ja SPEAKER FOR VEHICLE AND MOUNTING STRUCTURE OF THE SPEAKER

Shogo Iwaya *et al.*, assignors to Mitsubishi Jidosha Kogyo Kabushiki Kaisha
13 February 2007 (Class 381/389); filed in Japan 16 November 2001

Insert tabs **13** in slits **5** and rotate "circumferentially." If you actually
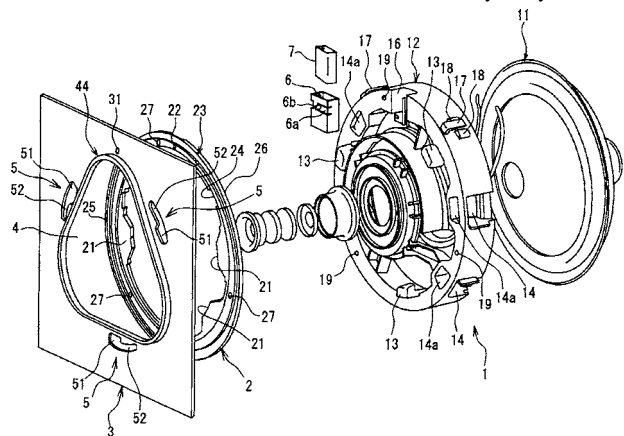
rotate counter–clockwise the speaker **11** is now installed in panel member **3**.—NAS

7,181,038

43.38.Ja ELECTRO-ACOUSTIC TRANSDUCER

Dean Menchaca Rivera and Christopher Sean Larson, assignors to KSC Industries Incorporated
20 February 2007 (Class 381/396); filed 10 March 2004

A robust loudspeaker design for various automobile applications is disclosed. The description includes how to center the voice-coil former **41** during assembly. Apparently, top portion **15** not only supports the magnetic
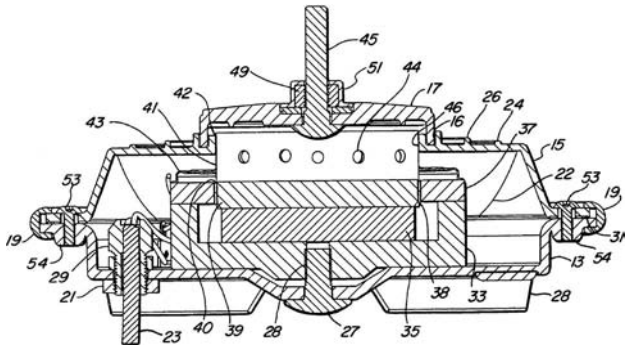
motor assembly parts (**33**, **35**, **37**, **38**, **39**, and **41**), and back cover **13**, and associated parts, but it also serves as the radiating portion of the device.—NAS

within the two cylindrical pole pieces. An additional "bucking" magnet **42** and shield plate **60** suppress stray magnetic leakage.—GLA

7,184,566

43.38.Ja BASS REFLEX TYPE SPEAKER DEVICE, MOUNTING STRUCTURE AND MOUNTING METHOD FOR SPEAKER DEVICE

Ichiro Tamura *et al.*, assignors to Sharp Kabushiki Kaisha
27 February 2007 (Class 381/388); filed in Japan 15 March 2002

Suppose you have a thin, panel-like structure that is, in fact, a visual display device. Within the thin depth you are concerned with providing satisfactory bass playback. You can use a bass reflex, or ported enclosure. If

you angle port **7** within the enclosure (the back portion **3** of which is shown), you may save some space.—NAS

7,200,241

43.38.Ja LOUDSPEAKER

Takanori Fukuyama and Shuhei Konishi, assignors to Matsushita
Electric Industrial Company, Limited
3 April 2007 (Class 381/407); filed 28 November 2002

In this elongated loudspeaker, rectangular voice coil **22** operates in a rectangular gap formed between center pole piece **25c** and outer pole pieces **25d**. Thus far, what has been described is known prior art. However, the patent argues that when this configuration is miniaturized, the clearance
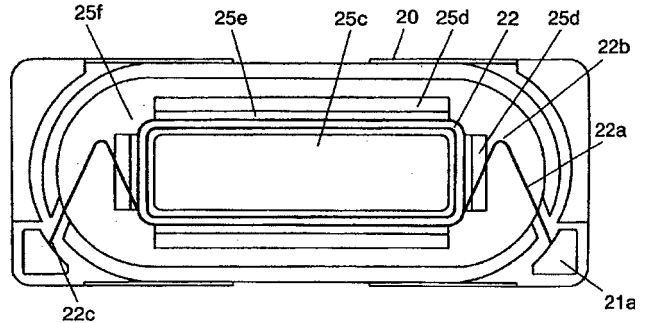
between the voice coil and outer frame **20** becomes so small that conventional connecting wires **22a** may stretch and break. The solution is to bend the wires at **22b**. This is probably the first time since the invention of the paper clip that a bent wire has been patented.—GLA

7,197,155

43.38.Ja MAGNET ASSEMBLY FOR LOUDSPEAKERS

Graham Bank, assignor to New Transducers Limited
27 March 2007 (Class 381/420); filed in United Kingdom 10 October 2002

In the magnetic assembly shown, inner pole piece **2** and outer pole piece **3** define voice coil gap **5**. A radially charged magnet **46** lies completely

7,201,252

43.38.Ja LOUDSPEAKER SYSTEMS

Stuart Michael Nevill, assignor to B&W Loudspeakers Limited
10 April 2007 (Class 181/151); filed in United Kingdom 21 September 2001

Many of B&W's commercial loudspeaker systems employ damped, tapered tubes to absorb the backwave energy of tweeters and mid-range speakers. This patent discloses a pancake arrangement in which multiple
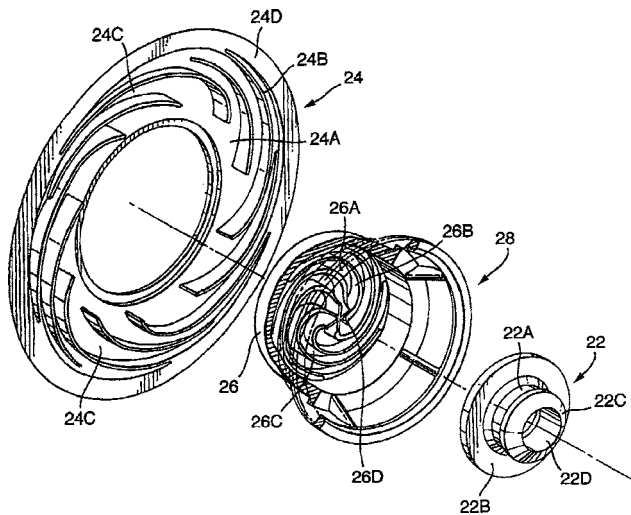
tubes **24c** are nested in a spiral configuration. The patent includes a number of refinements including controlled transmission between adjacent tubes to attenuate resonances and perforated cellular construction to provide controlled damping.—GLA

7,200,238

43.38.Lc ACOUSTIC SIGNAL LEVEL LIMITER

Ching Shyu *et al.*, assignors to Plantronics, Incorporated
3 April 2007 (Class 381/94.8); filed 28 September 2001

Symmetrical clipping is sometimes used to limit audio signal amplitude in small, low-power devices such as cellular telephones. This patent points out that if simple shunt circuits are driven into deep saturation, then the signal voltage can continue to rise and allow dangerous acoustic levels
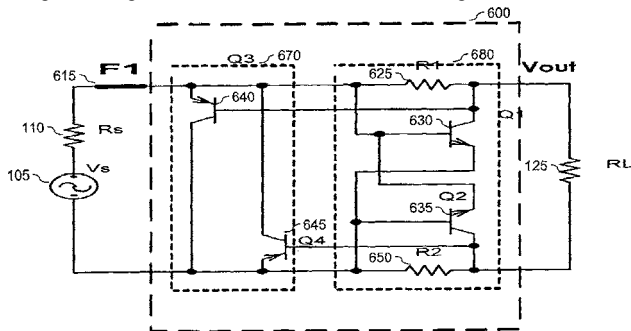
to be generated. (The issue of waveform distortion vs loudness is not discussed.) A multithreshold circuit is described in which at least two shunt circuits are progressively triggered. Finally, in the case of catastrophic overload, a fuse (not shown) is connected in series with the load.—GLA

7,181,025

43.38.Md ULTRASOUND BASED PARAMETRIC LOUDSPEAKER SYSTEM

Guido Kolano and Klaus Linhard, assignors to DaimlerChrysler AG
20 February 2007 (Class 381/77); filed in Germany 7 April 2001

Parametric loudspeakers based on demodulation of high-level ultrasound signals using arrays of piezoelectric transducers have been described by Yoneyama, Blackstock, and Pompei, among others. While these use
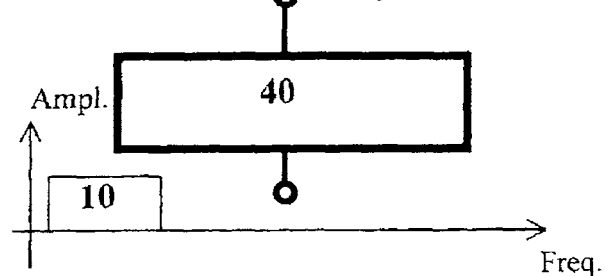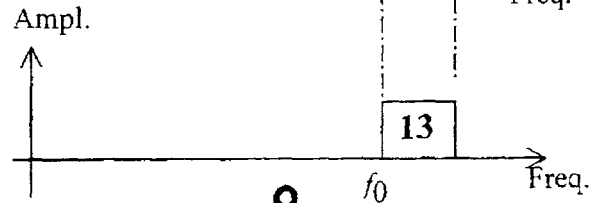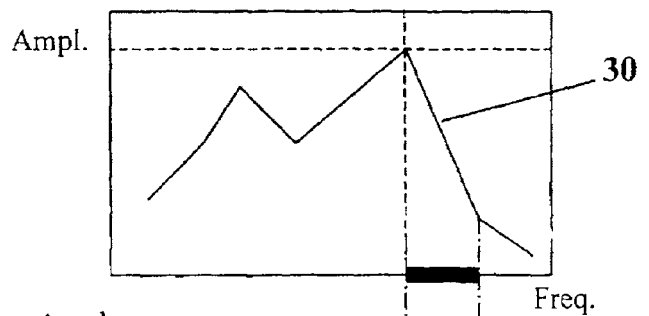
double-side-band amplitude modulation, the present patent describes a method in general terms that uses frequency modulation. This method is said to offer some advantages in matching the modulated signal to "particularly resonant transducers." The patent and its claims appear to be very general in their scope.—NAS

7,190,797

43.38.Si HEADSET WITH FOLDABLE NOISE CANCELING AND OMNIDIRECTIONAL DUAL-MODE BOOM

Tim Johnston and Kwangsee Allen Woo, assignors to Plantronics, Incorporated
13 March 2007 (Class 381/74); filed 18 June 2002

The sketch depicts a clip-on headset that might be used with a cellular telephone. Casing **12** houses a conventional earphone. Microphone boom **14** swivels so that it can be used in two configurations. When the boom is extended as shown, microphone **106** is near the user's mouth and operates as
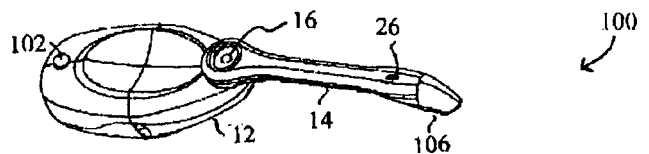
a noise cancelling microphone. When the boom is swiveled 180° and locked against the casing, the rear (noise cancelling) port is plugged and the microphone operates as a normal omnidirectional microphone:—GLA

7,193,346

43.38.Si MULTI-MODE VIBRATION GENERATOR FOR COMMUNICATION TERMINAL

Ju Ho Kim, assignor to Samsung Electro-Mechanics Company, Limited
20 March 2007 (Class 310/81); filed in Republic of Korea 24 September 2004

Now that every cell phone owner in the civilized world has his own customized ring tone melody, the next technological frontier appears to be customized vibration-mode sequences. The patent describes a dual assembly that houses two stacked vibration generators. "The second vibration generator generates a vibration of a direction or frequency different from a vibration or frequency of the first vibration generator, whereby various mode vibrations are generated."—GLA
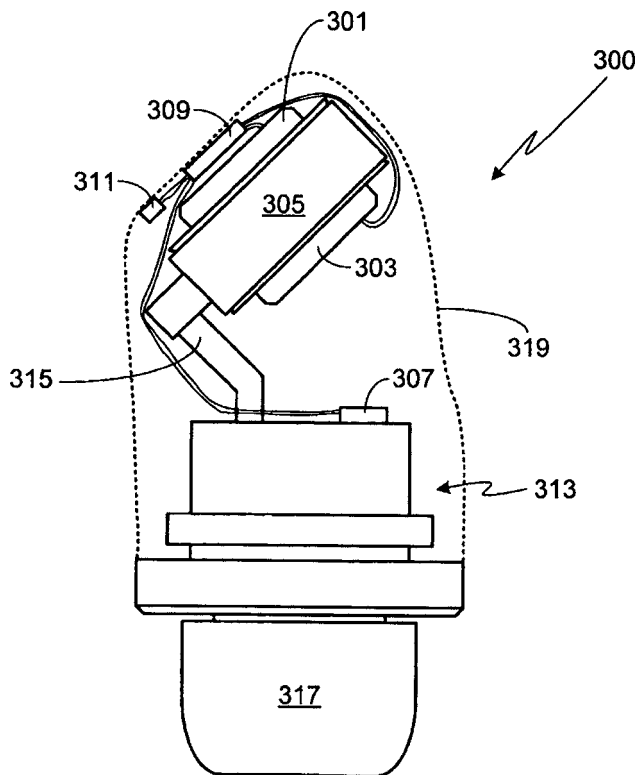
7,194,099

43.38.Si HANDHELD ELECTRONICS DEVICES WITH MULTIPLE USER SENSORY TRANSDUCERS AND METHODS

Ian Lewis, assignor to Motorola, Incorporated
20 March 2007 (Class 381/182); filed 10 June 2003

A number of existing patents describe dual-purpose transducers for use in cellular telephones and the like. In most cases, one voice coil drives a conventional loudspeaker diaphragm and the other drives a vibration generator. Well, why not install two separate devices? Better yet, why not patent

the basic concept of housing two transducers in a single handheld case? The only limitation is that both transducers must generate signal outputs. Otherwise, they can be the same or different, their response characteristics can be similar or dissimilar, and they can be driven from a single amplifier or two amplifiers.—GLA

7,194,102

43.38.Si IN-EAR MONITOR WITH HYBRID DUAL DIAPHRAGM AND SINGLE ARMATURE DESIGN

Jerry J. Harvey, assignor to Ultimate Ears, LLC
20 March 2007 (Class 381/328); filed 27 January 2005

This three-way, in-the-ear monitor houses two diaphragm type trans-

ducers and one balanced armature transducer to provide maximum sound pressure across the full range of audible frequencies.—GLA

7,197,140

43.38.Si SOUND BAFFLE FOR PORTABLE TELEPHONE HANDSET

V. Frank Asaro, San Diego, California
27 March 2007 (Class 379/433.03); filed 3 July 2003

Unlike a conventional telephone handset, today's cellular telephones are small, resulting in very little separation between the microphone and the earphone. One way to get the microphone closer to the user's mouth is to
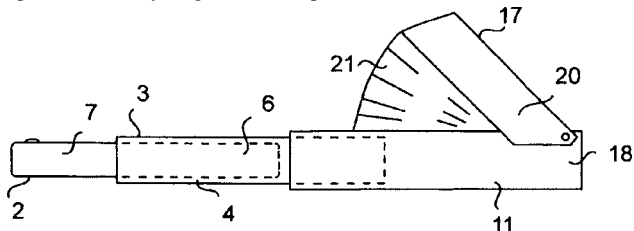
mount it on an extensible boom. This patent describes a different approach: A telescoping bellows-shaped waveguide not only conducts sound from the user's mouth back to the microphone but is also intended to "muffle his conversation to the surroundings for improved privacy."—GLA

7,197,149

43.38.Si CELLULAR PHONE

Ryota Mita and Akio Shinagawa, assignors to Hitachi, Limited
27 March 2007 (Class 381/98); filed in Japan 29 October 1999

In musical performances it is not uncommon to transpose a tune to a different key so that it will lie within the range of a particular instrument or vocalist. The frequency response of the miniature loudspeaker used in a typical cellular telephone is severely restricted, yet the user is allowed to

```
READ NOTE DATA    (S101)
        |
        v
DO THE DATA
INCLUDE A NOTE WHOSE    (S102)
SCALE IS LOWER        --YES-->  STEP UP THE WHOLE
THAN A3 ?                       DATA BY HALF STEP   (S103)
        |
       NO
        v
STORE THE NOTE DATA
IN THE MEMORY      (S104)
```
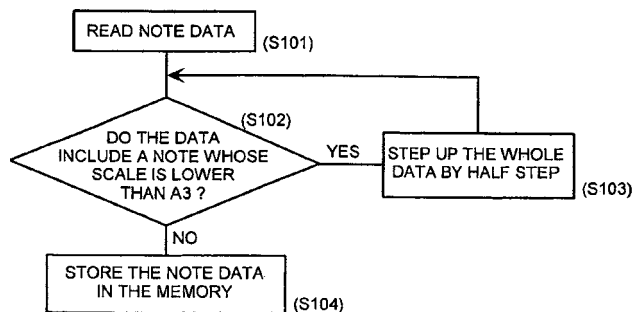
create custom ring melodies that can cover a much wider range. This patent describes an automated procedure that analyzes a stored melody, transposes it up or down in half-tone steps until it best fits within a specified frequency range, and then stores the transposed melody.—GLA

7,177,434

43.38.Tj STEPPED SOUND PRODUCING MODULE

Melchiore Tripoli, assignor to Sing-A-Tune Balloons, LLC
13 February 2007 (Class 381/190); filed 18 January 2002

Take piezoelectric module **20**, attach it to a "piezoelectric amplification device" **30**, which is then affixed to balloon **50** and you can produce
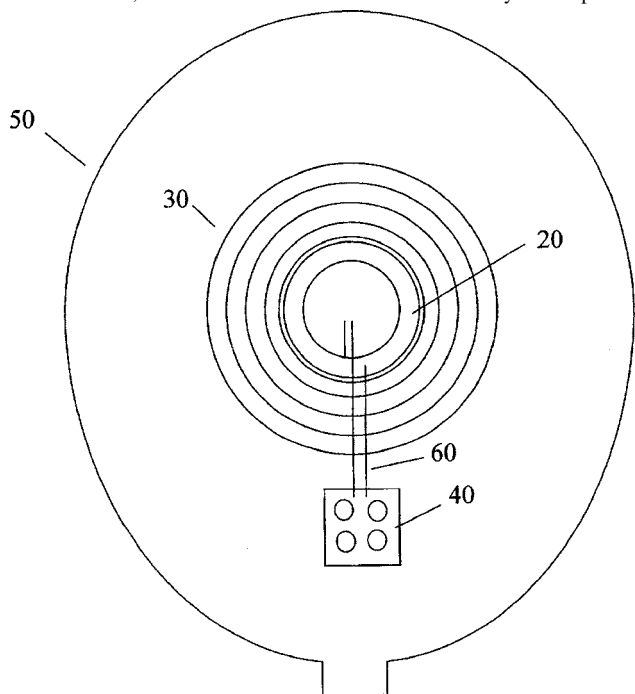
"hi-fidelity sound effects" by application of suitable signals to electric circuit **40**.—NAS

7,212,103

43.40.Le MONITOR SYSTEM, CENTRAL MONITOR APPARATUS, ON-VEHICLE MONITOR APPARATUS, MONITOR METHOD, MONITOR PROGRAM, COMPUTER-READABLE RECORDING MEDIUM CONTAINING THE SAME

Masayuki Oyagi et al., assignors to Omron Corporation
1 May 2007 (Class 340/429); filed in Japan 27 August 2001

This patent pertains to remotely monitored vehicle theft prevention systems, which employ vibration sensors to detect intrusions. In order to avoid false alarms that may result from earthquakes or nearby heavy machinery, several vehicles in the same general location (as determined from a

location determination system) are monitored simultaneously and concurrent intrusion alarms are discounted. Performance of the vibration sensors is also monitored in order to avoid false alarms from faulty sensors.—EEU

7,215,118

43.40.Le TRANSDUCER FOR GENERATING AND MEASURING TORSIONAL WAVES, AND APPARATUS AND METHOD FOR STRUCTURAL DIAGNOSIS USING THE SAME

Chan Il Park et al., assignors to Seoul National University Industry Foundation
8 May 2007 (Class 324/238); filed in Republic of Korea 16 February 2004

In this apparatus for the nondestructive testing of shafts or pipes, ferromagnetic strips **1** are placed around the test article **2** at an angle. A bias coil **6**, wound around an outer housing **5**, carries a direct current that increases the magnitude of the magnetic field that acts on the aforementioned
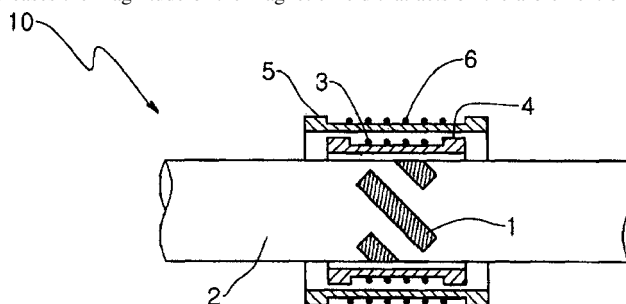
strips. When current is supplied to the coil **3**, which is wound around an inner housing **4**, torsional waves are induced in the test article. The waves reflected from the ends of the test article and from any cracks are detected at strips **1**, and identified by means of suitable display arrangements.—EEU

7,204,762

43.40.Tm SELF DAMPENING ROTARY SHAFT

Craig Andrew Campbell, assignor to GKN Driveline North America, Incorporated
17 April 2007 (Class 464/180); filed 24 March 2004

A hollow shaft, such as an automotive driveline shaft, is provided with damping by means of a rolled-up blanket that is placed into the shaft. The blanket may consist of layers of a flexible material, a heat-resistant material such as a ceramic, an adhesive, and a wire mesh.—EEU

7,210,356

43.40.Yq PHYSICAL AGENTS DIRECTIVE DOSIMETER SYSTEM

James R. Bernhagen, assignor to Caterpillar Incorporated
1 May 2007 (Class 73/661); filed 18 February 2005

An operator of a work machine (truck, excavator, etc.) is often exposed to significant vibration. In order to comply with the health and safety requirements limiting prolonged whole body vibration exposure, the patent describes a vibration "dosimeter system" comprising an accelerometer attached to the operator seat and a "controller" which compares the measured vibration level against predetermined exposure limits.—DMD

7,206,258

43.50.Cb DUAL RESPONSE ACOUSTICAL SENSOR SYSTEM

Stanley A. Fisher and Gideon Maidanik, assignors to United States of America as represented by the Secretary of the Navy
17 April 2007 (Class 367/141); filed 13 April 2005

This patent describes a method whereby both a (low-frequency) motion sensor and a (high-frequency) vibration sensor are embedded within the same compliant layer in order to facilitate optimal placement on ship hulls.—GFE

7,194,094

43.50.Fe SOUND MASKING SYSTEM

Thomas R. Horrall and John C. Heine, assignors to Acentech, Incorporated
20 March 2007 (Class 381/73.1); filed 24 October 2002

In describing a new speech masking system, this patent tiptoes carefully through a minefield of prior art. There appear to be two features that are unusual in this application. The first is that masking noise is conveyed to desired areas via direct sound coverage rather than by generally reflected
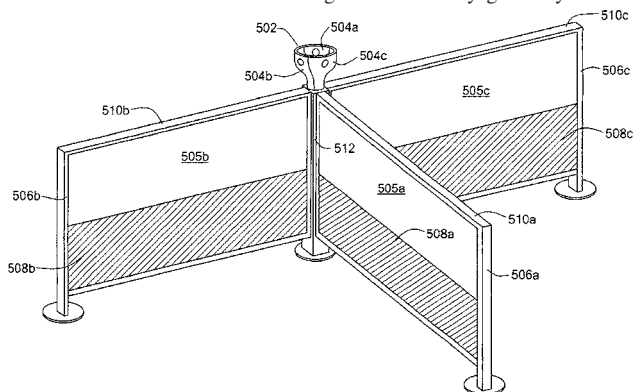
sound. The second is the use of a constricted aperture to broaden the high-frequency coverage of a loudspeaker. The patent claims also include a number of more questionable features, including an ordinary wireless remote control to adjust the operation of the system.—GLA

7,209,408

43.50.Rq DISTRIBUTED, SOFT-BODIED, TOWABLE, ACTIVE ACOUSTIC SYSTEM

Thomas R. Stottlemyer et al., assignors to United States of America represented by the Secretary of the Navy
24 April 2007 (Class 367/154); filed 7 October 2004

A soft tubed, active, acoustic, vertical array that can be recovered by conventional cable handling systems is described. Important issues such as transmit level and frequency are not discussed.—GFE

7,209,795

43.55.Cs METHOD OF SYNCHRONIZING THE PLAYBACK OF A DIGITAL AUDIO BROADCAST BY INSERTING A CONTROL TRACK PULSE

Gary E. Sullivan and Dusty D. Rector, assignors to Gateway Incorporated
24 April 2007 (Class 700/94); filed 23 April 2002

An obvious method to synchronize home surround speakers via a control track and time delays is described.—GFE

7,215,782

43.55.Ka APPARATUS AND METHOD FOR PRODUCING VIRTUAL ACOUSTIC SOUND

Jiashu Chen, assignor to Agere Systems Incorporated
8 May 2007 (Class 381/17); filed 23 January 2006

This patent describes a method said to produce realistic echoes for a virtual environment (e.g., games) based on a fast method of evaluating the room acoustics and head transfer function.—GFE

7,206,415

43.55.Mc AUTOMATED SOUND SYSTEM DESIGNING

Michael C. Monks et al., assignors to Bose Corporation
17 April 2007 (Class 381/58); filed 19 April 2002

This patent describes a method to automatically choose audio equipment for a room based on the room characteristics and the desired purpose.—GFE

7,216,853

43.55.Ti SOLID BARRIER SYSTEM

Michael D. Wall, Burton, Michigan
15 May 2007 (Class 256/24); filed 2 April 2004

This system includes a kit of parts that uses pre-existing cement anchored fence posts from a wire mesh cyclone fence. The parts wrap the post with elements that receive panels, and then the panels form a solid visual and acoustical screen. The panels are constructed of a lightweight one-piece molded material.—CJR

7,210,557

43.55.Vj LOW PROFILE ACOUSTIC FLOORING

John A. Phillips and Robert W. Hayes, assignors to ETS-Lindgren, L.P.
1 May 2007 (Class 181/207); filed 6 April 2004

The isolation element to support a floor panel for a sound isolation room is integrated into the floor itself so as to reduce overall height require-

ments. The system also allows for leveling of the floor after installation of the floor panel.—CJR
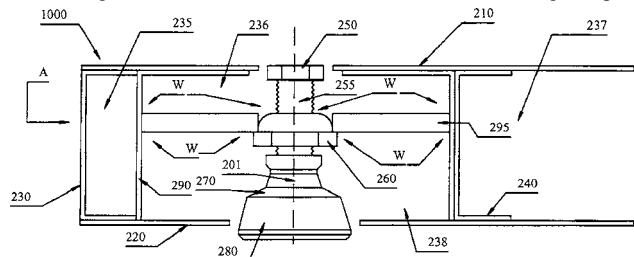
7,218,574

43.60.Dh HIGH RANGE RATE SIGNALING

Maurice D. Green, assignor to Teledyne Benthos, Incorporated
15 May 2007 (Class 367/134); filed 18 November 2004

A basic method is claimed to deal with the communication degradation caused by the relative velocity between transmitter and receiver. A simple preamble is transmitted before the communication signal to estimate range and remove Doppler spread.—GFE

7,054,452

43.60.Fg SIGNAL PROCESSING APPARATUS AND SIGNAL PROCESSING METHOD

Masakazu Ukita, assignor to Sony Corporation
30 May 2006 (Class 381/92); filed in Japan 24 August 2000

This is close to a traditional delay-and-add beamformer. The difference is that each microphone signal is processed through a filter bank before the cross-correlations are done. Correlations are then performed between bands of like frequency for all input signals. The rationale for this is that frequency bands which are dominated by noise will not show a strong peak at any
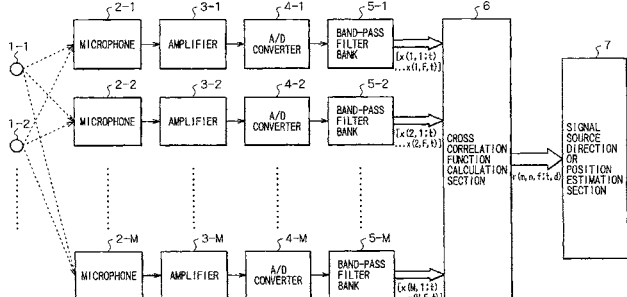
delay value within the correlator range. These bands may thus be excluded from the source direction estimate, improving the accuracy of an estimate based on those bands which do show a clear peak. There seems to be no mention of a preferred number of bands to be used. Presumably, even a few fairly wide bands would gain some of the expected advantage. The patent text includes detailed descriptions of various prior-art multimicrophone systems.—DLR

7,212,467

43.60.Fg SONAR LOCALIZATION

Peter Dobbins, assignor to BAE Systems (Land and Sea Systems) Limited
1 May 2007 (Class 367/138); filed in United Kingdom 5 October 2001

Sadly, it was only a matter of time before someone claimed to discover an acoustic divining rod. Here is another attempt to patent basic 2-D beamforming.—GFE

7,215,785

43.60.Fg PASSIVE SOUND TELEMETRY SYSTEM AND METHOD AND OPERATING TOY USING THE SAME

Sang Gyu Ju, Ahmsa-dong, Kangdong-ku, Seoul, Republic of Korea
8 May 2007 (Class 381/92); filed 3 February 2000

Acoustic signal processing in toys is apparently getting hot; however, the authors blithely lay a claim to the most basic concept in source localization; measuring phase delay.—GFE

7,212,466

43.60.Gk PRODUCING AMPLITUDE VALUES FOR CONTROLLING PIXEL ILLUMINATION ON A SONAR DISPLAY

Douglas James Wilson, assignor to Imagenex Technology Corporation
1 May 2007 (Class 367/68); filed 15 July 2004

"May" was the most frequent word in this dubious patent. It may describe a fish finder display that may be directed by user input.—GFE

7,218,240

43.66.Dc SYNTHETICALLY GENERATED SOUND CUES

Brian J. Tillotson, assignor to The Boeing Company
15 May 2007 (Oasis 340/692); filed 10 August 2004

The patent covers the concept of giving telemetry a spatial direction via acoustic masking. For example, a pilot might hear a radio transmission from the wingman, located off the right wing, solely in the right headphone.—GFE

7,198,596

43.66.Ts COUPLING DEVICE FOR A TWO-PART BONE-ANCHORED HEARING AID APPARATUS

Patrik Westerkull, assignor to P&B Research AB
3 April 2007 (Class 600/25); filed in Sweden 21 June 2001

Traditional bone-anchored hearing aid devices are coupled to the head with mechanically spring-actuated parts which over time are susceptible to mechanical wear and become loose. This patent recommends coupling via a
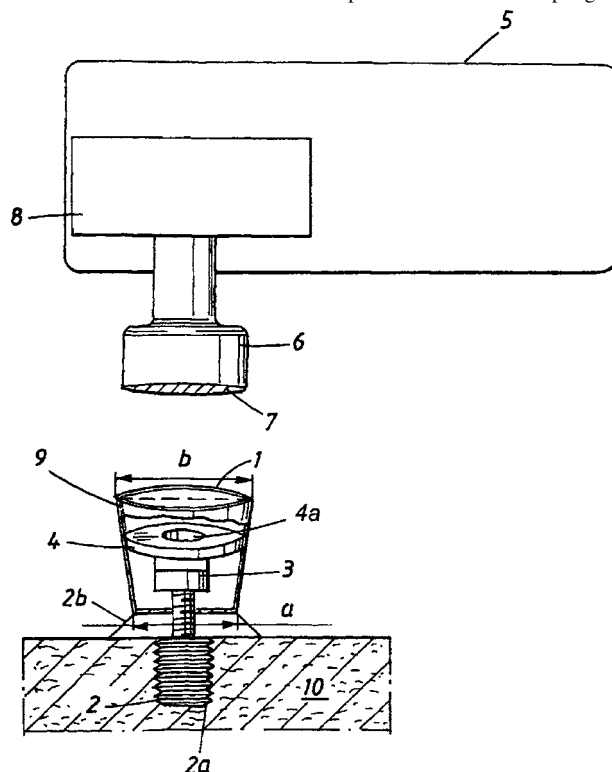
permanent magnet within the hearing aid device or the bone anchored portion. The magnet may be of sintered or ferromagnetic material and can be annular in shape and made integral with the skin-penetrating member and connecting screw.—DAP

7,204,799

43.66.Ts MICROPHONE OPTIMIZED FOR IMPLANT USE

Scott Allan Miller III and Bernd Waldmann, assignors to Otologics, LLC
17 April 2007 (Class 600/25); filed 5 November 2004

An implantable microphone uses many small diaphragms rather than a single diaphragm to reduce vibration sensitivity relative to acoustic sensi-

1864   J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Reviews of Acoustical Patents   1864
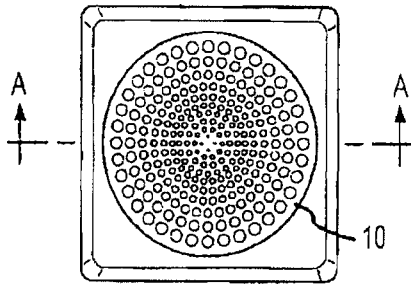
tivity. The array of small diaphragms may be created by pushing an irregular structure against a large membrane; additionally, an integrated diaphragm/support structure may also be utilized.—DAP

7,204,800

43.66.Ts IMPLANTABLE HEARING AID TRANSDUCER INTERFACE

James Roy Easter et al., assignors to Otologics, LLC
17 April 2007 (Class 600/25); filed 17 April 2006

In an implantable middle-ear hearing aid, the position of the ossicular chain relative to the implanted transducer is often changed due to changes in barometric pressure, tissue growth, and swallowing. To compensate for these physiological changes and to better isolate the microphone from
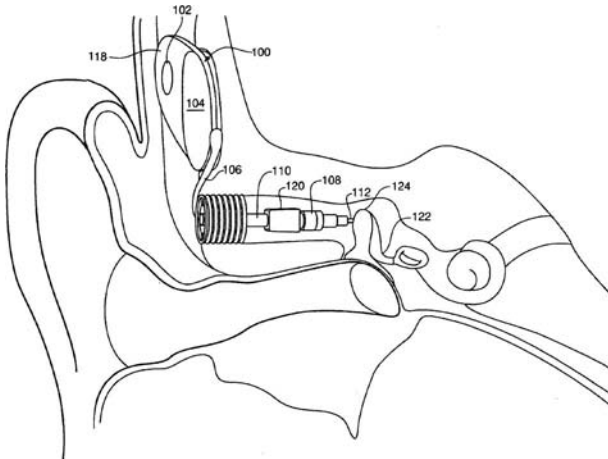
vibratory feedback produced by the transducer, a compliant interface, which may be a spring of biocompatible material, is provided for the implantable transducer. The spring rate of the interface is selected to produce a natural resonant frequency that is lower than the feedback frequency range.—DAP

7,205,918

43.66.Ts HEARING AID DEVICE WITH AN OUTPUT AMPLIFIER HAVING A SIGMA-DELTA MODULATOR

Torsten Niederdränk and Peter Nikles, assignors to Siemens Audiologische Technik GmbH
17 April 2007 (Class 341/143); filed in Germany 15 February 2005

The goal of this patent is to provide a hearing aid output stage that requires less battery current and has low internal circuit noise. A linear digital filter, which may be a delay element and an adder, is connected between the sigma-delta modulator and the output stage in order to generate three different voltage states at the output. The filter's frequency selectivity reduces the number of high-frequency edges and interference signals in the pulse-density-modulated output.—DAP

7,206,416

43.66.Ts SPEECH-BASED OPTIMIZATION OF DIGITAL HEARING DEVICES

Lee S. Krause et al., assignors to University of Florida Research Foundation, Incorporated
17 April 2007 (Class 381/60); filed 18 June 2004

Distinctive features of speech are used rather than conventional hearing tests or subjective responses for tuning a digital hearing aid or cochlear implant. Words and/or syllables representing one or more features of speech in a selectable language are played for a user and those misrecognized are analyzed with a confusion error matrix. The result of the error analysis generates appropriate hearing device intensity, frequency, and temporal parameter changes that are associated with the missed speech features.—DAP

7,206,423

43.66.Ts INTRABODY COMMUNICATION FOR A HEARING AID

Albert S. Feng et al., assignors to Board of Trustees of University of Illinois
17 April 2007 (Class 381/312); filed 10 May 2000

A personal area wireless network is created via conductive paths through the hearing aid wearer's skin. Modules housing acoustic input and

1865   J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Reviews of Acoustical Patents   1865

output components that have one- or two-way communication with each other via electrode pairs are located at the two ears of the wearer, forming a directional sensor array. These modules also communicate with a remote processing component that may be housed in a wrist watch or a cochlear implant.—DAP

7,206,424

43.66.Ts HEARING AID WITH TIME-VARYING PERFORMANCE

Mike K. Sacha, assignor to Starkey Laboratories, Incorporated
17 April 2007 (Class 381/315); filed 24 November 2004

In hearing aid fitting, new hearing aid wearers often experience an acclimatization period for getting accustomed to amplified sounds. Such hearing aid wearers have devices that are typically programmed initially to produce less acoustic gain and a narrower frequency response. Repeated visits to the dispensing professional allow for adjusting to higher gain and wider frequency response. To eliminate the need for these repeated visits, a hearing aid is programmed to successively select parameter sets automatically in a defined sequence. After preset numbers of power-up events or over several elapsed times, selected parameter sets would compensate for hearing loss in a gradually progressing way.—DAP

7,206,426

43.66.Ts MULTI-COIL COUPLING SYSTEM FOR HEARING AID APPLICATIONS

Steven D. Julstrom and Mead Killion, assignors to Etymotic Research, Incorporated
17 April 2007 (Class 381/331); filed 1 April 2004

The output of an array microphone or other audio source is transmitted from a behind-the-ear (BTE) hearing improvement device via wires or with inductive coupling to telecoils having different orientations in BTE and custom in-the-ear (ITE) hearing devices. To ensure sufficient magnetic
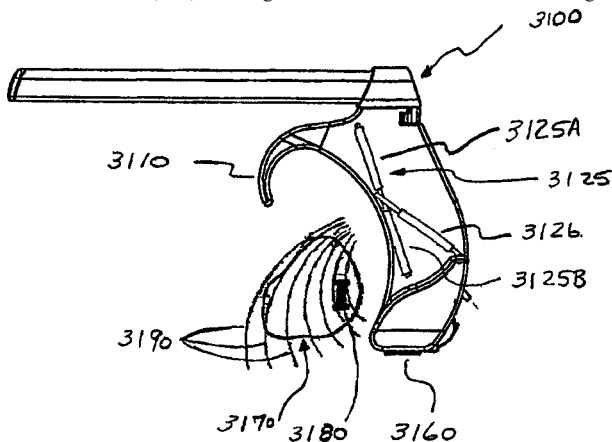
signal strength and uniformity of the induction coupling, the hearing improvement device contains two transmitting induction coils, one oriented for ITE and one for BTE hearing aids, respectively.—DAP

7,209,568

43.66.Ts HEARING AID HAVING AN ADJUSTABLE DIRECTIONAL CHARACTERISTIC, AND METHOD FOR ADJUSTMENT THEREOF

Georg-Erwin Arndt et al., assignors to Siemens Audiologische Technik GmbH
24 April 2007 (Class 381/313); filed in Germany 16 July 2003

The directional characteristic of a hearing aid having two or more omnidirectional microphones is varied by adjusting the time delay(s) in series with one or more of the microphone outputs as a function of the input
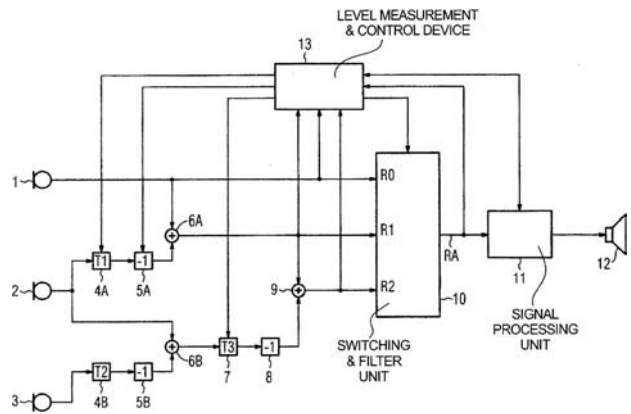
level into at least one of the microphones. The result is said to be improved signal-to-noise ratio while the hearing aid wearer is listening in quiet acoustical environments.—DAP

7,215,789

43.66.Ts DISPOSABLE EXTENDED WEAR CANAL HEARING DEVICE

Adnan Shennib and Richard C. Urso, assignors to InSound Medical, Incorporated
8 May 2007 (Class 381/328); filed 16 January 2002

A single-use hearing aid that fits completely into the ear canal consists of a lateral section with a microphone and battery, that is suspended without occluding in the cartilaginous part of the ear canal and a receiver section
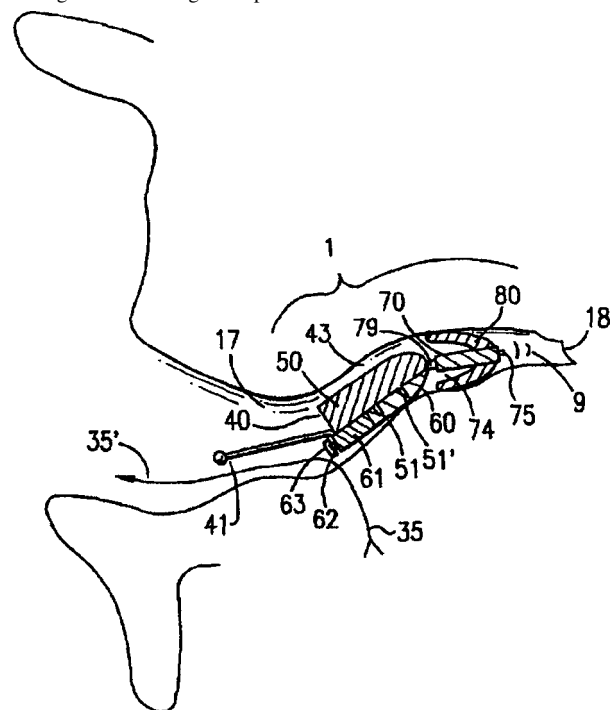
having a speaker and sealing retainer for the bony portion of the ear canal. The lateral section is flexible to permit movement in response to changes in ear canal shape and has an oval cross section with a cylindrical tapered shape along its longitudinal axis.—DAP

1866   J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Reviews of Acoustical Patents   1866

7,212,643

43.66.Ts REAL-EAR ZOOM HEARING DEVICE

Friedrich Bock, assignor to Phonak AG
1 May 2007 (Class 381/330); filed 10 February 2004

Intended for behind-the-ear hearing aids or hearing protectors, the object of this patent is to provide a transfer characteristic for sounds arriving from different directions that is similar to that of custom hearing aids placed in the ear canal. To simulate the pinna directivity effect, beamforming processing on the outputs of two microphones produces an omnidirectional response (independent of direction of impinging sound) at about 1 kHz and about 6 dB greater amplification for 45° than for 135° sound incidence at 5 kHz.—DAP

7,210,353

43.66.Yw MULTIMEDIA FEATURE FOR DIAGNOSTIC INSTRUMENTATION

Leroy Braun and Jack Foreman, assignors to Diagnostic Group, LLC
1 May 2007 (Class 73/585); filed 14 October 2003

A multimedia automatic audiometer includes means for computer-controlled switching between sound, visual graphics, or video signals stored in a computer and test tone outputs. The test subjects may select from a number of different languages through which testing is administered and test
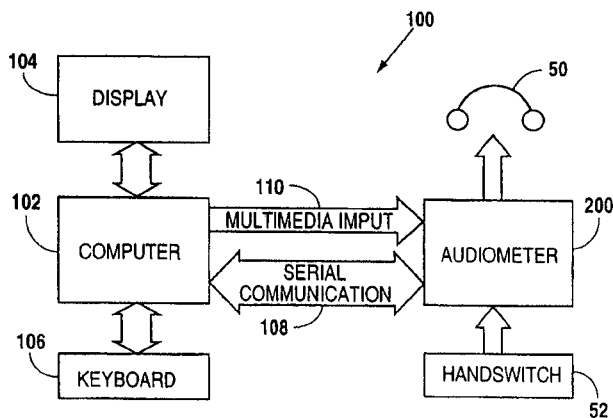
results may be graphically displayed. Several tests may be conducted simultaneously without human intervention. Corrective instructions are produced in earphones as a result of errors detected in subject's responses, followed by automatic switching to resume a test.—DAP

7,208,671

43.71.Gv SOUND DATA OUTPUT AND MANIPULATION USING HAPTIC FEEDBACK

Lonny L. Chu, assignor to Immersion Corporation
24 April 2007 (Class 84/609); filed 20 February 2004

A method is described to give musicians more intuitive feedback, via

the sense of touch, while they edit sounds on a computer.—GFE

7,054,806

43.72.Ja SPEECH SYNTHESIS APPARATUS USING PITCH MARKS, CONTROL METHOD THEREFOR, AND COMPUTER-READABLE MEMORY

Masayuki Yamada, assignor to Canon Kabushiki Kaisha
30 May 2006 (Class 704/207); filed in Japan 9 March 1998

This patent seems to be a response to some unspecified, but horribly inefficient, prior-art system for processing speech pitch intervals. It consists in its entirety of software to store and compare intervals from stored speech signals. The patent does not even include the pitch tracker, but assumes that pitch periods have been previously marked in the input speech signals. The claims include coverage of a "speech synthesis apparatus," but no such device is described in the text.—DLR

7,054,814

43.72.Ja METHOD AND APPARATUS OF SELECTING SEGMENTS FOR SPEECH SYNTHESIS BY WAY OF SPEECH SEGMENT RECOGNITION

Yasuo Okutani et al., assignors to Canon Kabushiki Kaisha
30 May 2006 (Class 704/256.4); filed in Japan 31 March 2000

The title of this patent seems to say that it is a speech synthesis system which uses a speech recognition system as part of the internal mechanism. The synthesis method at hand is that of concatenation of selected speech units to form the synthetic output. Such a system depends upon a large database of speech units in order to achieve high quality synthesis. Hidden Markov model (HMM)recognition technology based on Mel cepstrum analysis is put to use in constructing the speech unit database. A large selection of input speech is processed by the Mel-cepstral HMM recognizer in order to select those speech units most useful for building the synthesis database. Nothing more is said here about indexing methods used to organize the database.—DLR

7,054,815

43.72.Ja SPEECH SYNTHESIZING METHOD AND APPARATUS USING PROSODY CONTROL

Masayuki Yamada and Yasuhiro Komori, assignors to Canon Kabushiki Kaisha
30 May 2006 (Class 704/267); filed in Japan 31 March 2000

This system for speech synthesis performs a more or less typical type of phonetic unit concatenation in order to construct the synthetic speech signal. In order to improve the synthetic speech quality, the method disclosed here isolates individual pitch period segments from the selected phonetic units and uses a technique of grading the pitch period segments. By
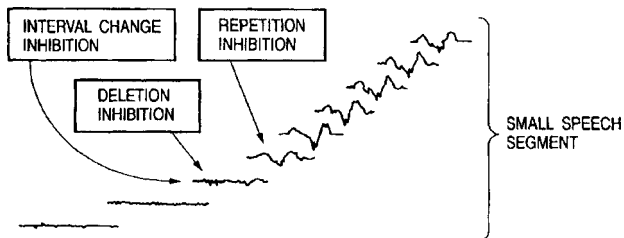
this process, some segments, particularly those which span intervals of significant waveform or spectral change, are marked as not subject to deletion, replication, or modification during the synthesis process. The remaining pitch period segments are then modified as needed in order to generate the desired prosodic pattern in the synthesized speech.—DLR

7,058,569

43.72.Ja FAST WAVEFORM SYNCHRONIZATION FOR CONCENTRATION AND TIME-SCALE MODIFICATION OF SPEECH

Geert Coorman and Bert Van Coile, assignors to Nuance Communications, Incorporated
6 June 2006 (Class 704/216); filed 14 September 2001

This method for concatenative speech synthesis proceeds by computing optimal join points for each speech unit in the concatenative database. These join points are determined such that two energy constraints are optimized, namely that the energy of both left and right segments be minimum, while also maximizing the energy in the crosscorrelation between the left
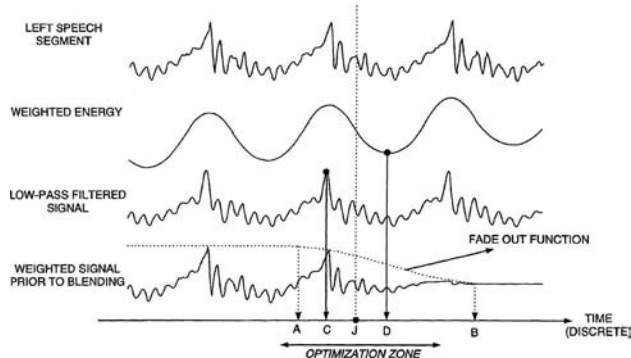
and right segments. Perceptually optimum results are achieved by considering the logarithm of the energy levels, rather than absolute energy. Considering the database as a whole and considering the range of pitch levels present in the database, it is possible to predetermine the optimum join points for each database segment independently. This allows all database segments to be premarked, thus reducing the time required for computing join points at synthesis time.—DLR

7,062,438

43.72.Ja SPEECH SYNTHESIS METHOD AND APPARATUS, PROGRAM, RECORDING MEDIUM AND ROBOT APPARATUS

Kenichiro Kobayashi et al., assignors to Sony Corporation
13 June 2006 (Class 704/260); filed in Japan 15 March 2002

The title of this patent states that a speech synthesizer is the subject, along with the related method, apparatus, etc. Only the word "robot" suggests that this synthesizer may be located in a somewhat special environment. The patent actually describes a humanoid robot, designed to be able to assist humans in a variety of ways, not only by cleaning the carpet, but by picking up and throwing a ball, getting up if it falls down, expressing feelings of joy, surprise, or sadness, and by being able to entertain its human
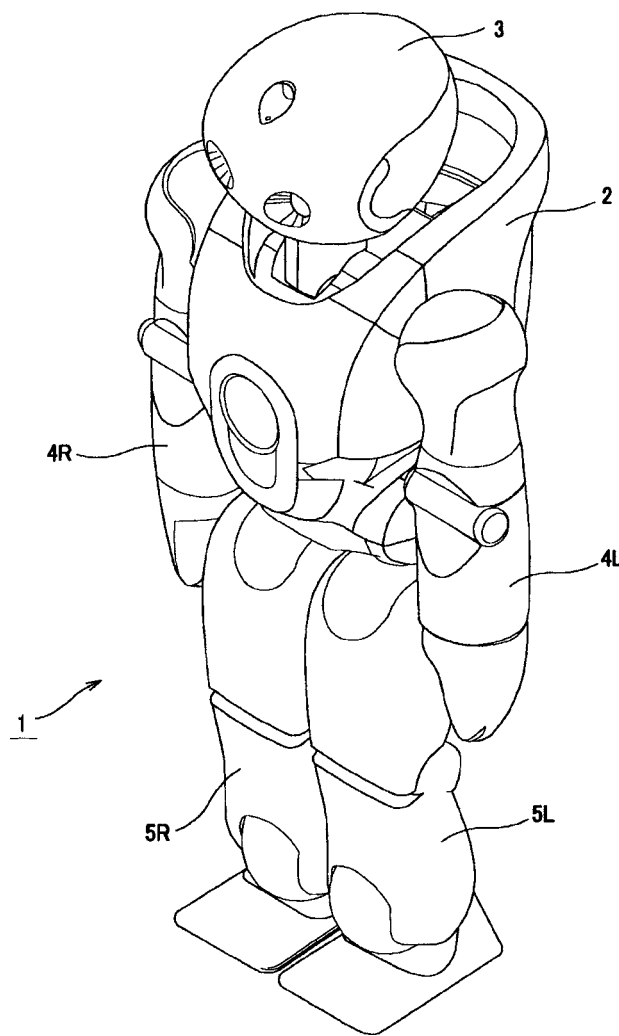
hosts by talking and signing, to mention just a few of the suggested capabilities. Most of the patent describes mechanisms for generating "close-to-human" quality speech, including a singing voice, following lyrics or music being played. A number of tables are presented with data describing how the speech prosodic structure would be generated. The claims deal exclusively with the vocal control mechanisms.—DLR

43.72.Ja SPEECH SYNTHESIS APPARATUS AND METHOD

Paul St John Brittan and Roger Cecil Ferry Tucker, assignors to Hewlett-Packard Development Company, L.P.
13 June 2006 (Class 704/260); filed in United Kingdom 4 June 2001

This patent describes a talking computer system, which includes a recognizer **11**, a dialog manager **7**, and a speech synthesizer **12**, but the emphasis here is on the synthesis system. More particularly, a mechanism is described by which the system can monitor its own speech output and evaluate the quality of the synthesized result. The core of the self-evaluation system seems to be that the speech is simply run through the recognition system and judged according to the usual recognition performance issues,
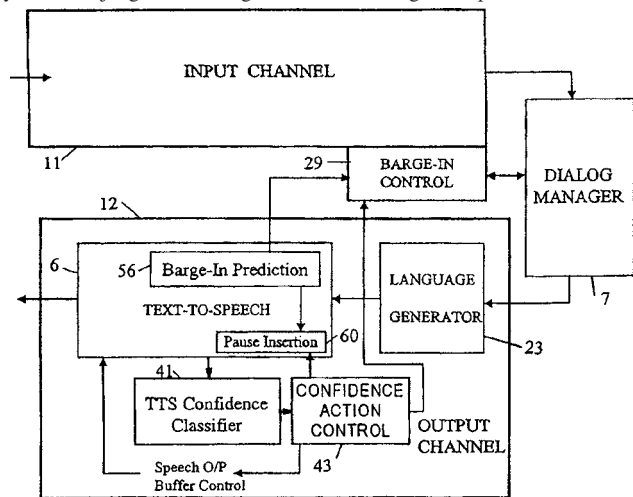
such as the percentage of correctly recognized phonemes, etc. But other characteristics are also monitored for their adequecy in expressing the state of the dialog control system, such as the appropriateness of the resulting prosodic patterns or the word-selection choices. The self-evaluation unit can then rephrase the utterance and resynthesize as needed to improve the evaluation.—DLR

7,209,882

43.72.Kb SYSTEM AND METHOD FOR TRIPHONE-BASED UNIT SELECTION FOR VISUAL SPEECH SYNTHESIS

Eric Cosatto et al., assignors to AT&T Corporation
24 April 2007 (Class 704/235); filed 10 May 2002

This patent describes a method to generate a video sequence (of a mouth moving) to an arbitrary tract of speech.—GFE

7,212,628

43.72.Kb ECHO CANCELLATION/SUPPRESSION AND DOUBLE-TALK DETECTION IN COMMUNICATION PATHS

Mirjana Popovic and Dieter Schultz, assignors to Mitel Networks Corporation
1 May 2007 (Class 379/406.08); filed in United Kingdom 31 January 2003

The patent describes a straight-forward method to suppress echoes via a level threshold.—GFE

43.72.Ne SPEECH INPUT SYSTEM, SPEECH PORTAL SERVER, AND SPEECH INPUT TERMINAL

Soshiro Kuzunuki et al., assignors to Hitachi, Limited
6 June 2006 (Class 704/270.1); filed in Japan 2 October 2001

What is covered here, by the single claim of this patent, is a run-of-the-mill speech recognizer, organized for use in a network environment. Such organization involves separating the speech input device, typically a microphone, from the recognition processor, using network protocols to transmit the audio signal. Details of a particular recognition architecture are not specified, but the claim does describe a grammar organization which allows the recognition results to be parsed into a command portion and an object text portion. The patent text also mentions a coding method by which letters may be extracted from recognized commands and used to build up a rapid-access command dictionary. Speech examples are presented in the patent in Japanese characters. The reader's Katakana skills will be helpful.—DLR

7,062,019

43.72.Ne METHOD OF PROVIDING SPEECH RECOGNITION FOR IVR AND VOICE MAIL SYSTEMS

Mike Krack, assignor to Avaya Technology Corporation
13 June 2006 (Class 379/88.04); filed 27 July 2001

This patent argues the case that the telephone system consists of a very large number of voice terminals, most of which can also easily transmit Touch-Tone™ (DTMF) signals, but can not easily receive such signals. In addition, this telephone system also includes a number of DTMF-capable voice messaging systems (VMSs), which can receive DTMF signals and transmit voice signals, but many of which cannot receive voice signals. What is thus needed for full interactive voice response (IVR) is a system, referred to here as an adjunct processor, or gateway, which has the capability of full interaction using voice signals and which can also transmit DTMF
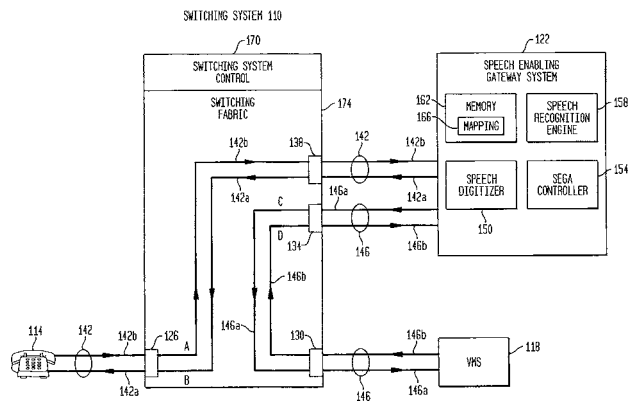
signals to be received by a VMS. Rather than using a party-line or conference-call type of telephone connection, the adjunct processor **122** would initially receive calls to be handled by the VMS **118** and would, in turn, call that system to process the customer's request. Details of the small-vocabulary speech recognition system are scant other than a listing of some command words which would be recognized and converted to DTMF tones.—DLR

7,054,813

43.72.Ne AUTOMATIC GENERATION OF EFFICIENT GRAMMAR FOR HEADING SELECTION

James R. Lewis *et al.*, **assignors to International Business Machines Corporation**
30 May 2006 (Class 704/251); filed 1 March 2002

The object of the patented speech recognition system is to use a "heading" to generate a small, efficient grammar for a specific speech recognition task based on a sample of the expected input speech. What is meant here by a heading is a "text word or phrase specifying the title, headline, content description or name of an associated book, chapter, subpart of a larger work, story, article, news item, electronic document, or the like." The set of words found in the heading is used to locate and install a small, efficient speech-recognition grammar suitable for the content of the item to be recognized. The patent also notes the possibility of using the heading information to help with the general task of document classification.—DLR

7,209,880

43.72.Ne SYSTEMS AND METHODS FOR DYNAMIC RE-CONFIGURABLE SPEECH RECOGNITION

Bojana Gajic *et al.*, **assignors to AT&T Corporation**
24 April 2007 (Class 704/231); filed 6 March 2002

Speech recognition as an input method for mobile devices is made more accurate by adapting automatically for individual transducer response characteristics and environmental background noise parameters. The system
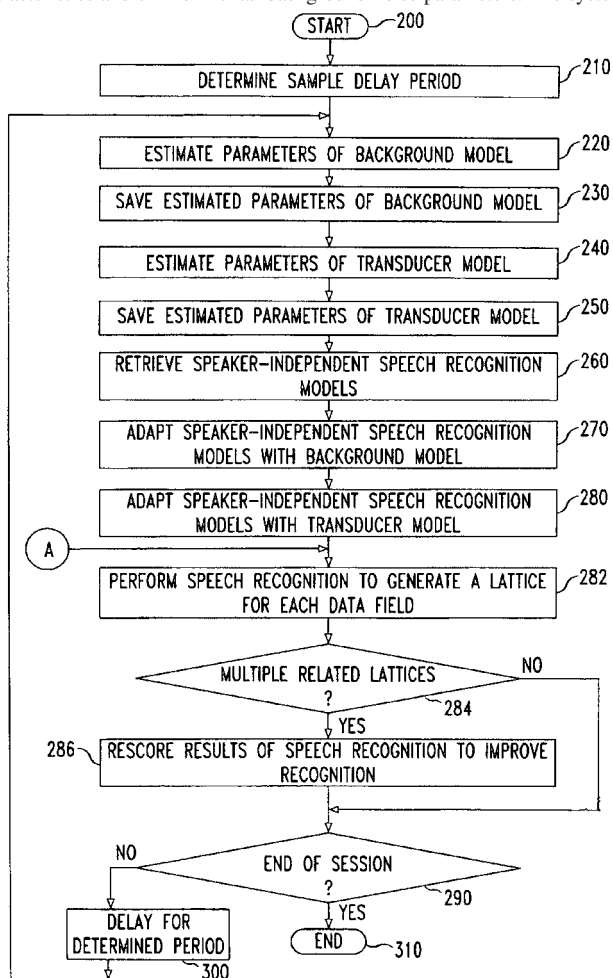
first determines the identity of a speaker via a unique user identifier code entered at the beginning of a session. This code is used to find linguistic constraints from previously recognized application fields. Word lattice interrelationships are determined by applying at least one language model to a single concatenated lattice resulting from speech utterances.—DAP

7,207,942

43.80.Vj ADAPTIVE GRATING LOBE SUPPRESSION IN ULTRASOUND IMAGING

Kutay F. Ustuner and Albert Gee, assignors to Siemens Medical Solutions USA, Incorporated
24 April 2007 (Class 600/443); filed 25 July 2003

The presence of grating lobe energy in received ultrasound echos is determined by processing the echos. Receive beamformation is then altered as a function of the grating lobe energy level to suppress grating lobes.—RCW

7,207,943

43.80.Vj SYNTHETIC ELEVATION APERTURE FOR ULTRASOUND SYSTEMS AND METHODS

Stephen R. Barnes *et al.*, **assignors to Siemens Medical Solutions USA, Incorporated**
24 April 2007 (Class 600/447); filed 24 March 2004

An aperture associated with a transducer in an ultrasonic imaging system is mechanically or electronically shifted and echos are obtained from different positions of the aperture. The echos are used to form a synthetic aperture for beamformation in the elevation dimension. Imaging is performed either without synthetic elevation aperture processing or by using synthetic elevation aperture processing with different elevation focusing and scanning parameters.—RCW

7,217,242

43.80.Vj ULTRASONIC METHOD FOR VISUALIZING BRACHYTHERAPHY SEEDS

Sheikh Kaisar Alam *et al.*, **assignors to Riverside Research Institute**
15 May 2007 (Class 600/439); filed 12 March 2003

Imaging of brachytherapy seeds is enhanced by illuminating them at their determined resonant frequency and using Doppler to detect their vibration, by determining their acoustic signature, and using cross-correlation to detect their presence, or by comparing images obtained before and after the application of stress.—RCW

43.80.Vj MULTIPLANAR ULTRASONIC VASCULAR IMAGING DEVICE, SYSTEM INCORPORATING SAME, METHOD OF USE AND PROTECTIVE SHEATH

Bradley J. Stringer *et al.*, **assignors to Inceptio Medical Technologies, L.C.**
8 May 2007 (Class 600/459); filed 21 June 2004

The transducer assembly in this system is used to obtain ultrasound images in two perpendicular planes. Markings on the assembly facilitate
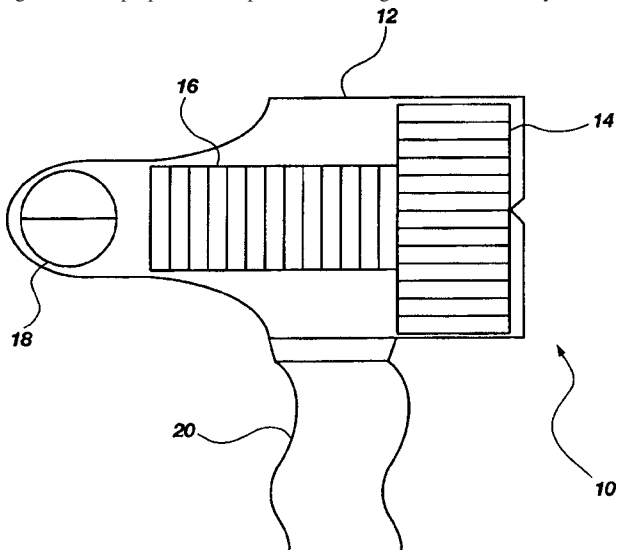
orientation of the transducer on the patient and guidance of a needle toward a target vessel during a cannulation procedure. The assembly includes a means to place, align, and secure it at a desired location on the skin of a patient.—RCW

43.80.Vj ULTRASOUND DIAGNOSIS APPARATUS

Hideki Takeuchi, assignor to Aloka Company, Limited
15 May 2007 (Class 600/447); filed in Japan 25 June 2003

A two-dimensional array is divided into subarrays for transmission and reception of ultrasound. Signals received from elements in the subarrays are
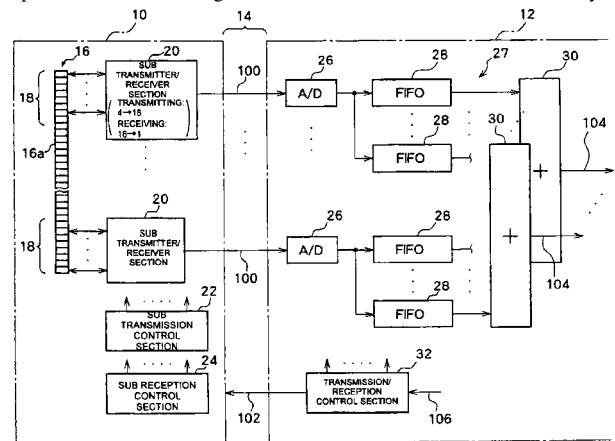
phase-adjusted and summed. The receive signal groups from the subarrays are also phase-adjusted and summed. Subarray elements are excited in groups during transmission.—RCW

LETTERS TO THE EDITOR

Existence of extraordinary zero-curvature slowness curve in anisotropic elastic media (L)

Litian Wang and Kent G. Ryne

Østfold University College, Engineering, 1757 Halden, Norway

Acoustic wave propagation in elastic media is characterized by the slowness surface. The slowness surface consists of three sheets associated with three modes of wave propagation and the two outer sheets can have zero-curvature locally. It is shown that the outmost sheet can admit extraordinary zero-curvature and the slowness curve can appear as a straight line locally. Using the perturbation method, the conditions for the extraordinary zero-curvature are derived analytically without violating the thermodynamic condition for elastic media. The results can be applied to crystals with higher symmetry and to the study of phonon focusing and surface waves. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2770546]

The slowness surface that characterizes acoustic wave propagation in elastic media plays a central role in understanding the anisotropic phenomena associated with bulk waves, phonon focusing, and surface waves.[1–4] As an important parameter, curvature of the slowness surface has been studied extensively in describing caustics in the phonon imaging where the caustics are related to parabolic lines on the slowness surface.[4] Since the slowness surface consists of three sheets associated with three-mode wave propagation and the two outer sheets can have zero-curvature, an interesting question can be raised as to whether the outermost sheet can have extraordinary zero-curvature. The extraordinary zero-curvature refers to the case where the slowness curve becomes straight locally in certain directions. The present work is to prove the existence of extraordinary zero-curvature in the vicinity of a twofold symmetry axis in monoclinic elastic media.

Wave propagation in anisotropic elastic media is governed by the Christoffel equation.[1–3] For a given propagation direction $\mathbf{k}$, the Christoffel equation is given by

$$\Gamma \mathbf{A} = \gamma \mathbf{A}, \tag{1}$$

where $\Gamma = k_i c_{ijkl} k_l$, $\gamma = \rho v^2$, and c_{ijkl} denotes the elastic stiffness tensor. Being a three-dimensional matrix, Γ yields three eigenvalues in terms of $\rho v^2(\mathbf{k})$, leaving three slowness surfaces $s(\mathbf{k}) = 1/v(\mathbf{k})$ associated with $|\mathbf{k}| = 1$. The innermost sheet describes the quasi-longitudinally polarized waves and is convex globally, while the two outersheets represent the quasi-transversely polarized waves and can be concave locally. The local concavity plays a central role in the study of phonon focusing and surface waves in anisotropic elastic media.[4–8]

Without losing generality, we consider a monoclinic medium with a twofold symmetry along the $\mathbf{e}_z$ axis. Monoclinic media are usually described by 13 elastic stiffness constants, and only 12 of them are independent. By properly choosing the coordinate system, we can transform the elastic tensor into the one with 12 elastic constants leaving $c_{45} = 0$. This is

done by setting $\mathbf{e}_x$ parallel to the polarization of the transverse mode defined by $\rho v^2 = c_{55}$ with $\mathbf{k} = \mathbf{e}_z$.[3] The following discussion is confined to the wave propagation along the $\mathbf{e}_z$ axis (the twofold symmetry axis), and the normal curvature with respect to $\mathbf{e}_z - \mathbf{e}_x$ plane. Defining the wave vector with $\mathbf{k} = \cos\theta \mathbf{e}_z + \sin\theta \mathbf{e}_x$, the matrix Γ can then be formulated by

$$\begin{pmatrix} c_{55} + \Delta_{15} \sin^2\theta & c_{16} \sin^2\theta & d_{13} \cos\theta \sin\theta \\ c_{16} \sin^2\theta & c_{44} + \Delta_{64} \sin^2\theta & c_{36} \cos\theta \sin\theta \\ d_{13} \cos\theta \sin\theta & c_{36} \cos\theta \sin\theta & c_{33} + \Delta_{53} \sin^2\theta \end{pmatrix}, \tag{2}$$

where $\Delta_{IJ} = c_{II} - c_{JJ}$ and $d_{13} = c_{13} + c_{55}$. By normalizing elastic constants against $c_{11}/\sqrt{\rho}$, the three eigenvalues γ_α give three slowness curves with $s_\alpha(\theta) = v_\alpha^{-1}(\theta) = \gamma_\alpha^{-1/2}(\theta)$.

It is well known that along the $\mathbf{e}_z$ axis the two outer slowness curves can have zero-curvature. However, whether the outermost slowness curve can have extraordinary zero-curvature has not been studied because its significance was not fully recognized. Geometrically, the extraordinary zero-curvature, if it is possible, would require vanishing first and/or second derivatives of the curvature. Because of the twofold symmetry about the $\mathbf{e}_z$ axis, the slowness curves are symmetric with respect to the $\mathbf{e}_z$ axis; it leaves $d\gamma/d\theta$ and $d^3\gamma/d\theta^3$ zero simultaneously. Since the slowness curves can be expressed in terms of one parameter θ, the normal curvature for the slowness curve and its second derivative in the vicinity of $\mathbf{e}_z$ axis can be defined by

$$k = -\gamma^{-1/2} \left(\gamma + \tfrac{1}{2}\gamma'' \right)\big|_{\theta=0}, \tag{3}$$

$$k'' = 4\gamma^{1/2} \left(\gamma - \tfrac{1}{8}\gamma^{(4)} \right)\big|_{\theta=0}. \tag{4}$$

The main difficulty is to derive general analytical expressions for the slowness curves. In the vicinity of the $\mathbf{e}_z$ axis, by rewriting the matrix Γ as $\Gamma^0 + V(\theta)$, where $\Gamma^0 = \text{diag}[c_{55}, c_{44}, c_{33}]$, we can obtain explicit expression for the slowness by using the perturbation method. We expanded the matrix $V(\theta)$ in the vicinity of the $\mathbf{e}_z$ axis ($\theta = 0$) and applied

the perturbation method up to the fourth order. The general expression for the fourth order correction on eigenvalues γ_n is given by

$$
\gamma_n^{(4)} = \sum_{m \neq n} \sum_{k \neq n} \sum_{p \neq n} \frac{V_{nm}V_{mk}V_{kp}V_{pn}}{(\gamma_n^0 - \gamma_m^0)(\gamma_n^0 - \gamma_k^0)(\gamma_n^0 - \gamma_p^0)}
$$
$$
- \sum_{m \neq n} \sum_{k \neq n} \left[\frac{V_{nm}V_{mk}V_{kn}V_{nn}}{(\gamma_n^0 - \gamma_m^0)^2(\gamma_n^0 - \gamma_k^0)} \right.
$$
$$
\left. + \frac{V_{nm}V_{mk}V_{kn}V_{nn}}{(\gamma_n^0 - \gamma_m^0)(\gamma_n^0 - \gamma_k^0)^2} \right]
$$
$$
- \sum_{m \neq n} \sum_{k \neq n} \left[\frac{|V_{mn}|^2|V_{nk}|^2}{(\gamma_n^0 - \gamma_m^0)^2(\gamma_n^0 - \gamma_k^0)} \right.
$$
$$
\left. + \frac{|V_{mn}|^2|V_{nk}|^2}{2(\gamma_n^0 - \gamma_m^0)(\gamma_n^0 - \gamma_k^0)^2} \right] + \sum_{m \neq n} \frac{|V_{mn}|^2|V_{nn}|^2}{(\gamma_n^0 - \gamma_m^0)^3}
$$
$$
+ \frac{1}{2} \left(\sum_{k \neq n} \frac{|V_{nk}|^2}{(\gamma_n^0 - \gamma_k^0)^2} \right) \left(\sum_{k \neq n} \frac{|V_{nk}|^2}{\gamma_n^0 - \gamma_k^0} \right).
$$

Note that the unperturbed eigenvalues are given by $\gamma_1^0 = c_{55}$, $\gamma_2^0 = c_{44}$, and $\gamma_3^0 = c_{33}$, and their corresponding eigenvectors are $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$, respectively. The eigenvalue γ_1 for the outermost slowness curve can then be explicitly expressed as

$$
\gamma_1 = c_{55} + V_{11} + \frac{V_{12}^2}{(\gamma_1^0 - \gamma_2^0)} + \frac{V_{13}^2}{(\gamma_1^0 - \gamma_3^0)} + \frac{V_{12}^2(V_{22} - V_{11})}{(\gamma_1^0 - \gamma_2^0)^2}
$$
$$
+ \frac{V_{13}^2(V_{33} - V_{11})}{(\gamma_1^0 - \gamma_3^0)^2} + \frac{2V_{12}V_{13}V_{23}}{(\gamma_1^0 - \gamma_2^0)(\gamma_1^0 - \gamma_3^0)}
$$
$$
+ \frac{V_{12}^2[(V_{22} - V_{11})^2 - V_{12}^2]}{(\gamma_1^0 - \gamma_2^0)^3} + \frac{V_{13}^2[(V_{33} - V_{11})^2 - V_{13}^2]}{(\gamma_1^0 - \gamma_3^0)^3}
$$
$$
+ \frac{V_{12}^2(V_{23}^2 - V_{13}^2) + 2V_{12}V_{13}V_{23}(V_{22} - V_{11})}{(\gamma_1^0 - \gamma_2^0)^2(\gamma_1^0 - \gamma_3^0)}
$$
$$
+ \frac{V_{13}^2(V_{23}^2 - V_{12}^2) + 2V_{12}V_{13}V_{23}(V_{33} - V_{33})}{(\gamma_1^0 - \gamma_2^0)(\gamma_1^0 - \gamma_3^0)^2}, \tag{5}
$$

where

$$
V_{11} = \Delta_{15}(\theta^2 - \theta^4/3), \quad V_{22} = \Delta_{64}(\theta^2 - \theta^4/3),
$$

$$
V_{33} = \Delta_{53}(\theta^2 - \theta^4/3), \quad V_{12} = c_{16}(\theta^2 - \theta^4/3),
$$

$$
V_{13} = d_{13}(\theta - 2\theta^3/3), \quad V_{23} = c_{36}(\theta - 2\theta^3/3).
$$

From (3) and (5), we can determine the normal curvature k_1 for the outermost slowness curve:

$$
k_1 = -(c_{11}\Delta_{35} - d_{13}^2)/(\Delta_{35}\sqrt{c_{55}}). \tag{6}
$$

When the normal curvature becomes zero, we would have a flat slowness curve in the vicinity of the $\mathbf{e}_z$ axis. The second derivative for the normal curvature k_1'' describes how the curvature changes in the neighborhood of the $\mathbf{e}_z$ axis. From (4) and (5), we have
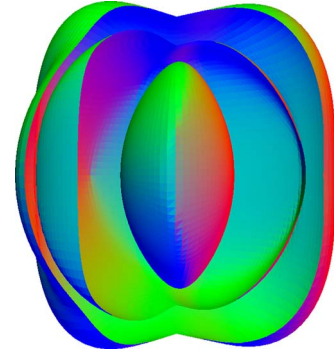
FIG. 1. (Color online) Slowness surface of an artificial monoclinic medium. Three-quarters of the slowness surface are illustrated with the viewpoint from $\langle \bar{1}10 \rangle$ direction. The outer sheet exhibits extraordinary zero-curvature in the neighborhood of $\mathbf{e}_z$ axis with respect to (100) and (010) planes.

$$
k_1'' = 4\sqrt{c_{55}}(\Delta_{35}^3\Delta_{45})^{-1}(-3\Delta_{45}d_{13}^4 + 3c_{36}^2\Delta_{35}d_{13}^2
$$
$$
- 4\Delta_{35}^2\Delta_{45}d_{13}^2 + 3(c_{11} + c_{33} - 2c_{55})\Delta_{35}\Delta_{45}d_{13}^2
$$
$$
- 6c_{16}c_{36}\Delta_{35}^2d_{13} + 3c_{16}^2\Delta_{35}^3 + c_{11}\Delta_{35}^3\Delta_{45}). \tag{7}
$$

When both k_1 and k_1'' vanish, the slowness curve will exhibit extraordinary zero-curvature and appears like a straight line locally in the neighborhood of the $\mathbf{e}_z$ axis.

Setting (6) and (7) to zero, we obtain the conditions for the existence of extraordinary zero-curvature in the vicinity of the twofold symmetry axis:

$$
c_{11}\Delta_{35} = d_{13}^2 \quad \text{and} \quad (c_{11}c_{36} - d_{13}c_{16})^2 = c_{11}^2c_{55}\Delta_{45}. \tag{8}
$$

A similar result can also be reached for the normal curvature along the $\mathbf{e}_z$ axis with respect to the $\mathbf{e}_z - \mathbf{e}_y$ plane. By defining the wave vector $\mathbf{k} = \cos\theta\mathbf{e}_z + \sin\theta\mathbf{e}_y$, the condition for the existence of extraordinary zero-curvature in the vicinity of $\theta = 0$ is given by

$$
c_{66}\Delta_{35} = c_{36}^2 \quad \text{and} \quad (c_{66}d_{23} - c_{26}c_{36})^2 = c_{66}^2c_{44}\Delta_{45}, \tag{9}
$$

where $d_{23} = c_{23} + c_{44}$.

Figure 1 illustrates the slowness sheets for an artificial monoclinic medium satisfying the conditions (8) and (9) and Fig. 2 illustrates two cross sections in the $\mathbf{e}_z - \mathbf{e}_x$ and $\mathbf{e}_z - \mathbf{e}_y$ planes, respectively. The medium is defined by the following elastic stiffness constants: $c_{11} = 1.0$, $c_{12} = 0.3$, $c_{13} = 0.25$, $c_{16} = 13/45$, $c_{22} = 0.8$, $c_{23} = 0.15$, $c_{26} = 0.3695$, $c_{33} = 0.4025$, c_{36}
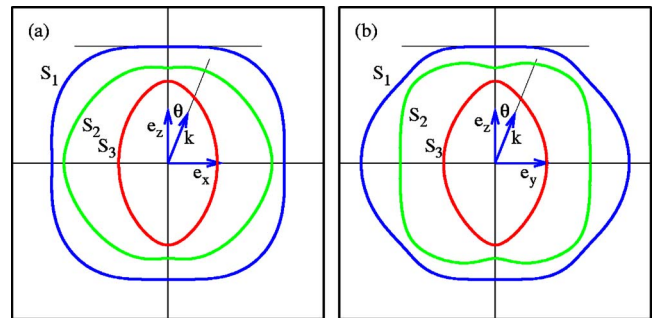
FIG. 2. (Color online) Slowness curves in (a) (010) plane and (b) (100) plane of the artificial monoclinic elastic medium. The local extraordinary zero-curvature on the outer slowness curve along $\langle 001 \rangle$ is marked with the straight lines.

$=0.27$, $c_{44}=0.298$, $c_{55}=0.2$, and $c_{66}=0.36$. It can be proven that the middle slowness curve can have zero-curvature $k_2=0$, but the second derivative k_2'' is always positive.

It should be emphasized that the conditions (8) and (9) do not contradict the basic requirement for thermodynamic stability. It can be shown that there is an eight-parameter family $(c_{11}, c_{12}, c_{13}, c_{22}, c_{23}, c_{36}, c_{44}, c_{55})$ of monoclinic elastic media that admit such an extraordinary zero-curvature slowness curve. The existence condition for such a slowness curve can also be satisfied in crystals with higher symmetry when a slowness curve is associated with a nonsymmetry cross section of the slowness surface.

The condition for the zero-curvature from (6) is consistent with the earlier studies.[1,2,7] The second derivative for the normal curvature k_1'' is of great significance in the theories of phonon focusing and surface waves. Geometrically, the slowness surfaces has degree of freedom 6, i.e., it can be intersected by a straight line at most at six points.[1] When a line contacts tangentially at the point with the extraordinary zero-curvature, the contact point can be regarded as six coinciding points on the surface. Such a point represents an extraordinary degeneracy in the Stroh formalism for two-dimensional elastodynamics.[8] In its neighborhood, one can expect that there exists a supersonic surface wave. In the phonon focusing, the extraordinary zero-curvature implies a boundary between concave and convex regions of the outer slowness sheet and can then be used to characterize the patterns for various types of caustics. It is worthwhile to mention that, in the phonon imaging,[4,7] Gaussian curvature, a product of two principal normal curvatures $K_1 \cdot K_2$, is usually used to describe the local geometry of slowness surfaces. Although the extraordinary zero-curvature discussed here refers to the transverse normal curvature with respect to two orthonormal planes, it may still be used to depict some characteristics of the local geometry of slowness surfaces because most of the characteristic points in the phonon focusing are related to the vanishing of the abovementioned transverse normal curvature.

To summarize, the slowness curve with extraordinary zero-curvature in the vicinity of the twofold symmetry axis in monoclinic elastic media is examined with the perturbation method up to the fourth order and is exemplified for the first time. More physical insights concerning the bulk/surface wave propagation and the phonon imaging remain to be explored.

[1] M. J. P. Musgrave, *Crystal Acoustics* (Holden-Day, San Francisco, 1970).
[2] B. A. Auld, *Acoustic Waves and Fields in Solids*, Vol. **1** (Wiley, New York, 1973).
[3] F. I. Fedorov, *Theory of Elastic Waves in Crystals* (Plenum, New York, 1968).
[4] J. P. Wolfe, *Imaging Phonons* (Cambridge U. P., Cambridge, 1998).
[5] A. L. Shuvalov and A. G. Every, "Curvature of acoustic slowness surface of anisotropic solids near symmetry axes," Phys. Rev. B **53**, 14906–14916 (1996).
[6] A. L. Shuvalov and A. G. Every, "Shape of the acoustic slowness surface of anisotropic solids near points of conical degeneracy," J. Acoust. Soc. Am. **101**, 2381–2383 (1997).
[7] A. L. Shuvalov and A. G. Every, "Transverse curvature of the acoustic slowness surface of anisotropic solids in crystal symmetry planes associated with phonon focusing cusps," J. Acoust. Soc. Am. **108**, 2107–2113 (2000).
[8] T. C. T. Ting, *Anisotropic Elasticity* (Oxford U. P., Oxford, 1996).

Further results for antiplane scattering by a thin strip (L)

Mihai Caleap and Christophe Aristégui[a]

Université Bordeaux 1, CNRS, UMR 5469, Laboratoire de Mécanique Physique, , Talence, F-33405 France

Yves C. Angel

University of Lyon, Lyon, F-69003, France, University Lyon 1, Lyon, F-69003, France
and INSERM, U556, Therapeutic Applications of Ultrasound Laboratory, Lyon, F-69424, France

A result, which does not appear to be available elsewhere, concerning the far-field scattering of antiplane waves by a single thin crack in an elastic solid is obtained in this letter. Specifically, the angular shape function is derived, with scattering coefficients expressed in terms of the crack-opening displacement. With this function, numerical values for effective speed and attenuation through a distribution of parallel cracks are obtained.
© *2007 Acoustical Society of America.* [DOI: 10.1121/1.2773994]

I. INTRODUCTION

Scattering of antiplane (SH) waves by thin cracks is re-examined in light of the theory proposed by Waterman and Truell (WT),[1] where scatterers are represented as equivalent line scatterers. In contrast, cracks have also been treated as finite-size scatterers.[2] In this letter, we compare the two approaches—that of line scatterers and that of finite-size scatterers. The WT approach uses the angular shape function corresponding to a single crack, which provides a measure of the wave motion scattered by the crack in the far field.

For this reason, we determine the exact analytical expression of the angular shape function in all directions. This result, to the best of our knowledge, is not available in analytical form in references on the subject.[3–10] In particular, we give exact expressions for the scattering coefficients in terms of the displacement discontinuity across the crack faces. The low-frequency limit of the far-field scattering is derived.

We determine next the effective wave number corresponding to a distribution of parallel cracks. We compare numerically the effective speed and attenuation with previous results,[2] where cracks are represented as finite-size scatterers. The results coincide for low crack concentration.

II. FORMULATION AND PRELIMINARIES

Consider a thin crack in a linearly elastic, homogeneous, and isotropic solid subjected to SH waves, as shown in Fig. 1. The crack has width $2a$ and infinite length in the y_2 direction. In the (y_1, y_3) plane, it is regarded as a line of discontinuity across which the component u_2 of the displacement is discontinuous. For normal incidence of a SH wave traveling in the y_3 direction, the scattered displacement u_2 is antisymmetrical with respect to the plane $y_3 = 0$. We consider a steady-state situation. In the following, we omit the time-factor $\exp(-i\omega t)$, which is common to the incident wave and to the displacement and stress components in the solid. The symbol ω stands for angular frequency. We define the unknown function b in the interval $(-a, a)$ by[7]

$$u_2(y_1, 0^+) = \begin{cases} \displaystyle\int_{y_1}^{a} b(v)dv, & |y_1| < a \\ 0, & |y_1| \geq a, \end{cases} \quad (1)$$

together with

$$\int_{-a}^{a} b(v)dv = 0. \quad (2)$$

The function b may be interpreted as a *dislocation density* across the crack faces.

Let c_T be the speed of transverse waves in the solid. The scattered displacement u_2 satisfies the Helmholtz equation in the solid with wave number $k = \omega/c_T$ and the boundary condition

$$u_2(y_1, y_3) = 0, \quad |y_1| \geq a, \quad y_3 = 0. \quad (3)$$

By taking the Fourier transform of the Helmholtz equation for u_2 with respect to the y_1 variable, solving the resulting ordinary differential equation, and using the boundary condition (3), together with Eqs. (1) and (2), one finds that the function b is a solution of the following singular integral equation:[7]

$$\int_{-a}^{a} b(v)\left[\frac{1}{v - y_1} + S(v - y_1)\right]dv = i\pi u_0 k, \quad |y_1| < a, \quad (4)$$

where u_0 is an amplitude factor for the displacement of the time-harmonic incident wave, and i is the imaginary unit. The function S is given by

$$S(y) = \int_{0}^{\infty} \left(\frac{\beta}{\xi} - 1\right)\sin(\xi y)d\xi, \quad (5)$$

with β defined by $\beta^2 = \xi^2 - k^2$ and Im $\beta \leq 0$.

We can express the steady-state scattered displacement u_2 at any point in the solid by using Green's representation formula. One finds that[7]

[a]Electronic mail: c.aristegui@lmp.u-bordeaux1.fr
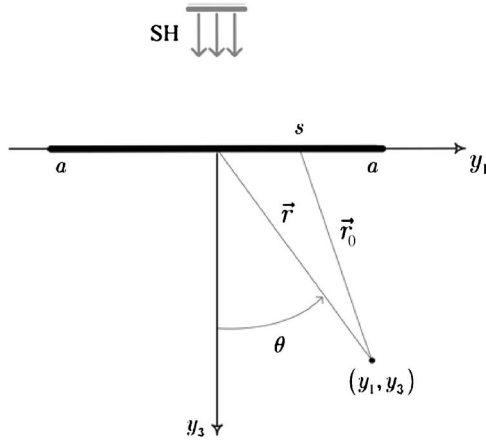
FIG. 1. Incident SH wave impinging normally on a crack.

$$u_2(y_1, y_3) = \frac{i}{2} \operatorname{sgn}(y_3) k y_3 \int_{-a}^{a} \frac{u_2(s, 0^+)}{r_0} H_1^{(1)}(k r_0) ds, \quad (6)$$

where r_0 is defined by $r_0^2 = (y_1 - s)^2 + y_3^2$. In Eq. (6), sgn denotes the sign function, and $H_1^{(1)}$ is the Hankel function of first order.

III. FAR-FIELD SCATTERING

Let (r, θ) denote the position of a point in the half-plane $y_3 > 0$, as shown in Fig. 1. The distance r and the angle θ are defined, respectively, by $r^2 = y_1^2 + y_3^2$ and $\cos \theta = y_3/r$. Thus, for $r \gg s$, one has

$$r_0 \simeq r\left(1 - \frac{s}{r} \sin \theta\right) + O(r^{-1}). \quad (7)$$

Then, one infers for $r \gg a$ that

$$H_1^{(1)}(k r_0) \simeq \sqrt{\frac{2}{\pi k r}} \exp\left(i\left(kr - ks \sin \theta - \frac{3\pi}{4}\right)\right)$$
$$\simeq H_1^{(1)}(kr) e^{-iks \sin \theta}. \quad (8)$$

It follows from Eqs. (6)–(8) that

$$u_2(r, \theta) = \frac{i}{2} k \cos \theta H_1^{(1)}(kr) \int_{-a}^{a} u_2(s, 0^+) e^{-iks \sin \theta} ds \quad (9)$$

$$(r \gg a, \ 0 \leq \theta < 2\pi).$$

Observe that the displacement $u_2(s, 0^+)$ is an even function of s for normally incident SH waves. Thus, it follows from Eq. (9) that $u_2(r, \theta) = u_2(r, -\theta)$, with θ oriented as in Fig. 1. Substituting Eqs. (1) and (8) into Eq. (9), then interchanging the integration order and using Eq. (2) in the resulting expression, one finds that the far-field scattered displacement (9) can be written in the equivalent form

$$u_2(r, \theta) = u_0 \sqrt{\frac{2\pi}{kr}} \exp\left(i\left(kr + \frac{\pi}{4}\right)\right) f(\theta), \quad (10)$$

where $f(\theta) = f(-\theta)$ and $f(\pi - \theta) = -f(\theta)$. In Eq. (10), the *angular shape function* f is defined by

$$f(\theta) = \frac{k \cos \theta}{2 i \pi} B(\theta), \quad (11)$$

where

$$B(\theta) = \frac{i}{k u_0 \sin \theta} \int_{-a}^{a} b(v) e^{-ikv \sin \theta} dv \quad (12)$$

$$(\theta \neq n\pi, \ n = 0, 1, 2, \dots).$$

In the neighborhood of $\theta = 0$ or π, using a Taylor series expansion of the exponential function in Eq. (12), together with Eq. (2), one finds that

$$f(0) = -f(\pi) = \frac{k}{2 i \pi} B(0), \quad (13)$$

with

$$B(0) = \frac{1}{u_0} \int_{-a}^{a} v b(v) dv. \quad (14)$$

We write the angular shape function in the well-known form

$$f(\theta) = \frac{1}{i\pi} \sum_{n=0}^{\infty} \varepsilon_n C_n \cos n\theta, \quad (15)$$

where $\varepsilon_0 = 1$ and $\varepsilon_n = 2$, for $n \geq 1$, and C_n represent the *scattering coefficients* of the crack. Inverting the formula (15), one has

$$C_p = i \int_0^{\pi} f(\theta) \cos(p\theta) d\theta, \quad p = 0, 1, 2, \dots. \quad (16)$$

After some lengthy but straightforward calculations, using the relations (1), (2), (11), and (12), one finds that the scattering coefficients are given for integer values $p \geq 0$ by

$$C_{2p} = 0, \quad (17)$$

$$C_{2p+1} = \frac{1}{2u_0} \int_{-a}^{a} \frac{u_2(\alpha, 0^+)}{\alpha} \left[(2p+1) J_{2p+1}(k\alpha) \right.$$
$$\left. + i(2p+1) E_{2p+1}(k\alpha) - i\frac{2}{\pi} \right] d\alpha. \quad (18)$$

In Eq. (18), J_p and E_p are, respectively, Bessel and Weber functions.[11] Since the term C_0 is that of isotropic scattering, we infer from Eq. (17) that a crack cannot scatter isotropically.

IV. LOW-FREQUENCY APPROXIMATION

We present here the Rayleigh limit approximation of the angular shape function. In Eq. (4), we introduce the notations $\tilde{\omega} = ka$ and $\tilde{S}(x) = aS(ax)$. In the limit as $\tilde{\omega}$ approaches zero, the function $\tilde{S}$ takes the asymptotic form[2]

$$\tilde{S}(x) \simeq \frac{1}{2} x \tilde{\omega}^2 \ln \tilde{\omega} - \frac{i\pi}{4} x \tilde{\omega}^2 \quad (\tilde{\omega} \to 0). \quad (19)$$

The function b that solves Eq. (4), together with Eq. (19), can be found in closed form.[12] Then, substituting b into Eqs.

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Caleap *et al.*: Letters to the Editor   1877

(12) and (14), and leaving aside real terms of $O(\tilde{\omega}^5 \ln \tilde{\omega})$ and imaginary terms of $O(\tilde{\omega}^3 \ln \tilde{\omega})$, one finds that

$$B(0) = B(\theta) \simeq a\left(\frac{\mathrm{i}\pi\tilde{\omega}}{2} - \frac{1}{16}\pi^2\tilde{\omega}^3 \right) \quad (\tilde{\omega} \to 0). \qquad (20)$$

Observe that the limit (20) is independent of the angle θ. Substituting Eq. (20) into Eq. (11), one finds that

$$f(\theta) = \frac{2}{\mathrm{i}\pi}C_1 \cos\theta \quad (\tilde{\omega} \to 0), \qquad (21)$$

where the scattering coefficient C_1 is given by

$$C_1 = -\frac{\pi^2}{64}\tilde{\omega}^4 + \mathrm{i}\frac{\pi}{8}\tilde{\omega}^2. \qquad (22)$$

The other coefficients $C_3, C_5, \dots$ are of order smaller than C_1 in the limit as $\tilde{\omega} \to 0$.

V. MULTIPLY CRACKED ELASTIC SOLID

Consider a homogeneous, isotropic, and linearly elastic solid containing a random and uniform distribution of identical and parallel cracks of width $2a$, such as that of Fig. 1. The cracks lie in the (y_1, y_2) plane. Let n be the number of cracks per unit area. The effective wave number K of the SH coherent wave propagating in the y_3 direction can be calculated as in WT.[1,13] The result is

$$K^2 = k^2\left(1 + n\frac{2\pi}{k^2}f(0) \right)^2 - k^2\left(n\frac{2\pi}{k^2}f(\pi) \right)^2, \qquad (23)$$

where $k = \omega/c_T$ is the wave number in the uncracked solid. One infers from Eqs. (13) and (23) that

$$K^2 = k(k - 2\mathrm{i}nB(0)). \qquad (24)$$

We recall that in Angel and Koba[2] a complex-valued wave number is determined for SH coherent waves in a solid containing parallel cracks. These authors write boundary conditions for the average exciting field on each finite-size crack. The cracks are not represented as line scatterers. Their formula [Eq. (48)], with our notations, is

$$K^2 = \frac{k^3}{k + 2\mathrm{i}nB(0)}. \qquad (25)$$

Let $\varepsilon = na^2$ be the dimensionless crack density. Then, in the limit as ε approaches zero, the expressions (24) and (25) are identical to within terms of $O(\varepsilon^2)$.

The attenuation α and the speed c of the SH coherent wave are defined by $K(\omega) = \omega/c(\omega) + \mathrm{i}\alpha(\omega)$. In the limit as $\tilde{\omega} \to 0$, using Eq. (20), one finds that

$$\alpha a = \frac{\varepsilon\pi^2}{16\sqrt{1 + \varepsilon\pi}}\tilde{\omega}^3 + O(\tilde{\omega}^5) \quad \text{and} \qquad (26)$$

$$c = \frac{c_T}{\sqrt{1 + \varepsilon\pi}} + O(\tilde{\omega}^2 \ln \tilde{\omega}).$$

We observe that the dominant terms in Eq. (26) are identical to within terms of $O(\varepsilon^2)$, respectively, to those in Eq. (83) of Angel and Koba.[2]
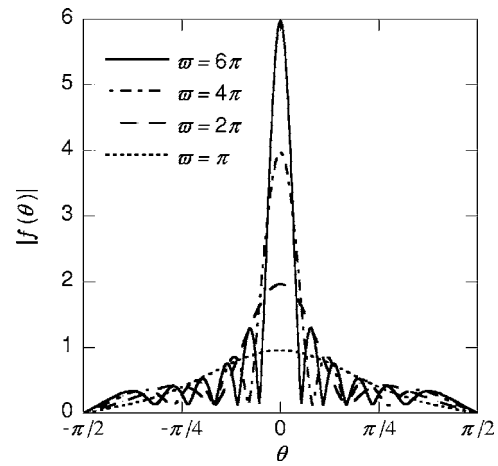
FIG. 2. Modulus $|f(\theta)|$ of the angular shape function vs the observation angle θ for values of the dimensionless frequency $\tilde{\omega} = \pi$, 2π, 4π, and 6π. The number of lobes increases with frequency.

VI. RESULTS AND ANALYSIS

The dislocation density b in Eqs. (12) and (14) is obtained by solving numerically the singular integral equation (4) with the auxiliary equation (2).[14] For four values of $\tilde{\omega}$, the modulus $|f(\theta)|$ of the angular shape function is shown in Fig. 2 versus the angle θ; it is symmetrical about $\theta = 0$, and vanishes at $\theta = \pm\pi/2$, as expected. It takes the same values on either side of the crack ($y_3 > 0$ or $y_3 < 0$). The maximum of $|f(\theta)|$ occurs at $\theta = 0$, for all values of $\tilde{\omega}$. More sidelobes appear as $\tilde{\omega}$ increases.

For normal incidence, the attenuation α and the speed c of the SH coherent wave in a solid containing identical parallel cracks are calculated by using Eq. (24). Results are presented in Figs. 3 and 4, respectively, for αa and c/c_T versus the dimensionless frequency $\tilde{\omega}$. The crack density takes values $\varepsilon = 0.01$, 0.03, 0.05. In these figures, earlier predictions are also displayed with triangular symbols.[2] We find very good agreement between the two approaches for $\varepsilon = 0.01$ and 0.03, as expected from the discussion in the preceding section.
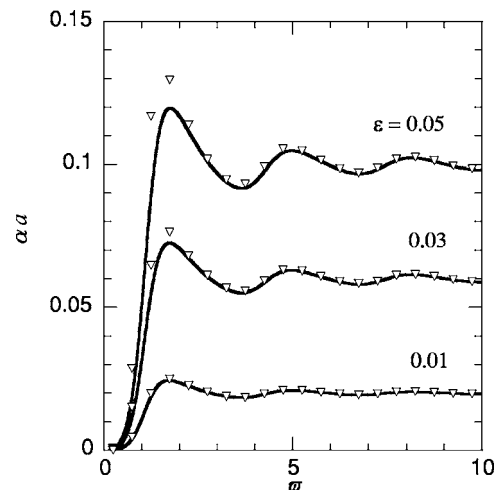
FIG. 3. Attenuation vs the dimensionless frequency for crack densities $\varepsilon = 0.01$, 0.03, and 0.05.
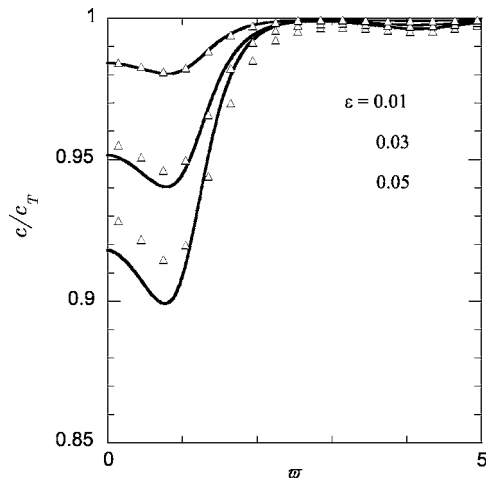
FIG. 4. Wave speed vs the dimensionless frequency for crack densities ε =0.01, 0.03, and 0.05.

VII. CONCLUSION

We have considered the scattering of normally incident SH waves by a two-dimensional line crack of finite width embedded in an unbounded elastic solid. We have derived analytical expressions for the angular shape function and the scattering coefficients. This has been used to evaluate the coherent scattering of SH waves in elastic solids containing multiple cracks.

[1] P. C. Waterman and R. Truell, "Multiple scattering of waves," J. Math. Phys. **2**, 512–537 (1961).

[2] Y. C. Angel and Y. K. Koba, "Complex-valued wavenumber, reflection and transmission in an elastic solid containing a cracked slab region," Int. J. Solids Struct. **35**, 573–592 (1998).

[3] J. F. Loeber and G. C. Sih, "Diffraction of antiplane shear waves by a finite crack," J. Acoust. Soc. Am. **44**, 90–98 (1968).

[4] R. L. Weaver and Y.-H. Pao, "Application of the transition matrix to a ribbon-shaped scatterer," J. Acoust. Soc. Am. **66**, 1199–1206 (1979).

[5] P. M. van den Berg, "Transition matrix in acoustic scattering by a strip," J. Acoust. Soc. Am. **70**, 615–619 (1981).

[6] J. D. Achenbach, A. K. Gautesen, and H. McMaken, *Ray Methods for Waves in Elastic Solids: With Applications to Scattering by Cracks* (Pitman Advanced Publishing, Boston, 1982).

[7] Y. C. Angel, "On the reduction of elastodynamic crack problems to singular integral equations," Int. J. Eng. Sci. **26**, 757–764 (1988).

[8] J. Y. Huang, "Interaction of SH-waves with a finite crack in a half-space," Eng. Fract. Mech. **51**, 217–224 (1995).

[9] C. Zhang and D. Gross, *On Wave Propagation in Elastic Solids with Cracks* (Computational Mechanics, Southampton, UK, 1998).

[10] J. F. Sánchez-Sesma and U. Iturrarán-Viveros, "Scattering and diffraction of SH waves by a finite crack: an analytical solution," Geophys. J. Int. **145**, 749–758 (2001).

[11] I. S. Grashteyn and I. W. Ryzhik, *Tables of Integrals, Series and Products*, 5th ed. (Academic, New York, 1994).

[12] N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Groningen-Holland, 1953).

[13] R.-B. Yang and A. K. Mal, "Multiple scattering of elastic waves in a fiber-reinforced composite," J. Mech. Phys. Solids **42**, 1945–1968 (1994).

[14] F. Erdogan and G. D. Gupta, "On the numerical solutions of singular integral equations," Q. Appl. Math. **30**, 525–534 (1972).

A further test of the linearity of temporal summation in forward masking (L)

Christopher J. Plack[a)] and Samuele Carcagno
Department of Psychology, Lancaster University, Lancaster, LA1 4YF, United Kingdom

Andrew J. Oxenham
Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, Minnesota 55455

An experiment tested the hypothesis that the masking effects of two nonoverlapping forward maskers are summed linearly over time. First, the levels of individual noise maskers required to mask a brief 4-kHz signal presented at 10-, 20-, 30-, or 40-dB sensation level (SL) were found. The hypothesis predicts that a combination of the first masker presented at the level required to mask the 10-dB SL signal and the second masker presented at the level required to mask the 20-dB SL signal, should produce the same amount of masking as the converse situation (i.e., the first masker presented at the level required to mask the 20-dB SL signal and the second masker presented at the level required to mask the 10-dB SL signal), and similarly for the 30- and 40-dB SL signals. The results were consistent with the predictions. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2775287]

I. INTRODUCTION

Forward masking refers to the decrease in the detectability of a signal as a result of prior stimulation by a masker. Forward masking has been shown to be highly nonlinear in listeners with normal hearing. For example, a given increase in the level of a forward masker produces a much smaller increase in the signal level at threshold, for signal levels less than about 30 dB SPL (Jesteadt *et al.*, 1982; Moore and Glasberg, 1983; Munson and Gardner, 1950). However, it has been demonstrated that these nonlinear effects can be simulated by a model that incorporates a compressive nonlinearity, representing the response of the basilar membrane (BM), prior to a *linear* leaky integrator or temporal window (Plack and Oxenham, 1998; Plack *et al.*, 2002). The output of the temporal window is a linear weighted sum over time of a quantity proportional to the square of BM velocity.

One of the predictions of linear summation is that the contribution of a stimulus to masking should be unaffected by the presence of stimuli before and/or after; the internal representations of two stimuli separated in time should be *independent*. In a recent study the prediction was tested using two nonoverlapping forward maskers (Plack *et al.*, 2006). It was demonstrated that the contribution of the second masker to masking was unaffected by the first masker, even when the first masker in the sequence rendered the second masker inaudible. The present experiment is a further test of this prediction. Consider a masking situation with two consecutive forward maskers, M1 and M2. The levels of the forward maskers are chosen so that in the first condition M1 has the level required to mask a signal with level L_s dB and M2 has the level required to mask a signal with level L_s+x

dB, and in the second condition M1 has the level required to mask a signal with level L_s+x dB and M2 has the level required to mask a signal with level L_s dB. We assume that for a given signal level the masking effect (in terms of the temporal window model, this is the output of the window in response to the masker at the time of signal presentation) is constant at threshold, irrespective of the time of presentation or duration of the masker, for example. Hence, the combined effect of the maskers is the same in both conditions: In each case the total masking effect is the sum of that required to mask a signal with level L_s dB and that required to mask a signal with level L_s+x dB. Because the masking effect is the same, the linear-summation hypothesis predicts that the signal thresholds in the two conditions should be *identical*. If, however, there is a nonlinear interaction between the maskers, for example if the first masker reduces the effectiveness of the second when the first masker is higher in level, the prediction will not hold. The present study therefore provides an empirical test of this prediction, which arises from the hypothesis of linear summation, to further test the validity of the temporal window model as an account of auditory temporal masking.

II. METHOD

A. Stimuli

The signal was a 4000-Hz pure tone. The first masker in the sequence (M1) was a Gaussian noise, bandpass filtered between 2800 and 5600 Hz (3-dB cutoffs, 90 dB/octave). The second masker (M2) was a Gaussian noise, bandpass filtered between 3400 and 4800 Hz. The signal had a total duration of 4 ms, which consisted of 2-ms raised-cosine onset and offset ramps (no steady state). Quoted levels are peak equivalent sound pressure levels. M1 had a total duration of 200 ms, including 2-ms onset and offset ramps and a 196-ms

---

a)Author to whom correspondence should be addressed. Electronic mail: c.plack@lancaster.ac.uk

TABLE I. The absolute thresholds for the signal and the results of Phase 1 of the experiment. The masker levels at threshold are given in decibel spectrum level for each sensation level of the signal. Standard errors are given in parentheses.

	Absolute threshold for signal (dB SPL)	10 dB SL		20 dB SL		30 dB SL		40 dB SL	
		M1	M2	M1	M2	M1	M2	M1	M2
L1	21.9 (0.5)	−2.3 (0.5)	7.6 (0.9)	13.3 (1.5)	24.4 (2.0)	28.9 (1.4)	34.3 (1.5)	47.3 (1.1)	44.0 (0.9)
L2	12.9 (0.4)	−11.0 (0.3)	14.8 (1.1)	5.1 (1.6)	31.8 (0.6)	19.4 (1.1)	40.5 (0.4)	30.0 (0.5)	47.8 (0.8)
L3	11.1 (1.1)	−14.9 (0.8)	8.9 (4.0)	0.3 (1.1)	25.8 (1.1)	15.3 (1.7)	34.3 (2.0)	25.0 (0.7)	39.2 (0.7)
L4	14.3 (0.4)	−16.1 (0.4)	−6.3 (2.2)	−6.7 (0.7)	11.9 (2.9)	−1.9 (3.0)	24.5 (3.3)	11.7 (0.9)	36.8 (3.0)
Mean	15.0 (2.4)	−11.1 (3.1)	6.3 (4.5)	3.0 (4.2)	23.5 (4.2)	15.4 (6.4)	33.4 (3.3)	28.5 (7.4)	41.9 (2.5)

steady-state portion. M2 had a total duration of 6 ms, including 2-ms onset and offset ramps and a 2-ms steady-state portion. The end of M1 coincided with the start of M2. The silent interval between the end of M2 and the start of the signal was 4 ms.

The experiment was controlled by custom-made software from a PC workstation located outside a double-walled sound-attenuating booth. All stimuli were generated digitally with 32-bit resolution and were output by an RME Digi96/8 PAD 24-bit soundcard set at a clocking rate of 48 kHz. The headphone output of the soundcard was fed via a patch panel in the sound booth wall to Sennheiser HD580 headphones without filtering or amplification. Stimuli were presented to the right ear. Each listener sat in the booth and decisions were recorded via a computer keyboard. Listeners viewed a computer monitor through a window in the sound booth. Lights on the monitor display flashed on and off concurrently with each stimulus presentation and provided feedback at the end of each trial.

B. Procedure

Four normally hearing listeners (ages 23–29) took part in the experiments. Listeners were given at least 2 h training on the conditions before data collection. The procedure was similar to that described in previous articles (Plack and O'Hanlon, 2003; Plack et al., 2006). For both absolute and masked thresholds, a three-interval forced-choice procedure was used with an interstimulus interval of 300 ms. Threshold was determined using a two-up one-down (Phase 1) or a two-down one-up (absolute thresholds and Phase 2) adaptive procedure that tracked the 70.7 percent correct point on the psychometric function (Levitt, 1971). The step size was 4 dB for the first four turnpoints, and was reduced to 2 dB for 12 subsequent turnpoints. The mean of the last 12 turnpoints was taken as the threshold estimate for each block of trials. At least four estimates were made for each condition and the results averaged.

First, absolute thresholds for the signal were measured for each listener. The main experiment was then conducted in two phases:

(i) Phase 1. The level of the signal was fixed at 10-, 20-, 30-, or 40-dB sensation level (SL). In each case, the

levels of M1 and M2 required to mask the signal when presented individually were determined. For M1, these maskers are designated M1(10), M1(20), M1(30), and M1(40) and similarly for M2.[1]

(ii) Phase 2. The threshold for the signal was determined in the presence of M1, M2, and M1+M2 (M1 and M2 combined). M1 and M2 were presented at the levels determined in Phase 1. The combined masker conditions were M1(10)+M2(20), M1(20)+M2(10), M1(30)+M2(40), M1(40)+M2(30).

III. RESULTS AND ANALYSIS

A. Results

The absolute thresholds for the signal, and the results of Phase 1, are presented in Table I. Despite the similarity in absolute thresholds, there is considerable variability in the masker levels required to mask the signal, particularly at the higher signal levels. However, this is not uncommon in situations in which the signal and/or masker are at sound levels greater than about 30 dB SPL and are therefore subject to strong compression (e.g. Plack and Drga, 2003; Plack and O'Hanlon, 2003).

The results of Phase 2 are presented in Fig. 1. The thresholds for the single masker conditions should be roughly equal to the signal levels presented in Phase 1. Although this is generally the case at low levels, the mean data show that the thresholds are less than expected at the higher levels (30 and 40 dB SL correspond to mean signal levels of 45 and 55 dB SPL; see Table I). The combined masker thresholds are, in general, considerably higher than the highest single masker threshold. With the exceptions of listeners L1 and L4 in the M1(20)/M2(10) condition, "excess" masking is observed in all cases. This is indicative of a compressive system (Oxenham and Moore, 1995). A repeated-measures analysis of variance (ANOVA) [overall level $(10/20, 30/40) \times$ masker sequence (low M1/high M2, high M1/low M2)] conducted on the combined masker thresholds revealed a highly significant effect of overall level [$F(1,3)=1090, p<0.0005$] but no effect of masker sequence [$F(1,3)=0.17, p=0.71$] and no interaction [$F(1,3)=2.24, p=0.23$]. The results from L1 and L4 in the M1(20)/M2(10) condition were somewhat anomalous, in that less excess
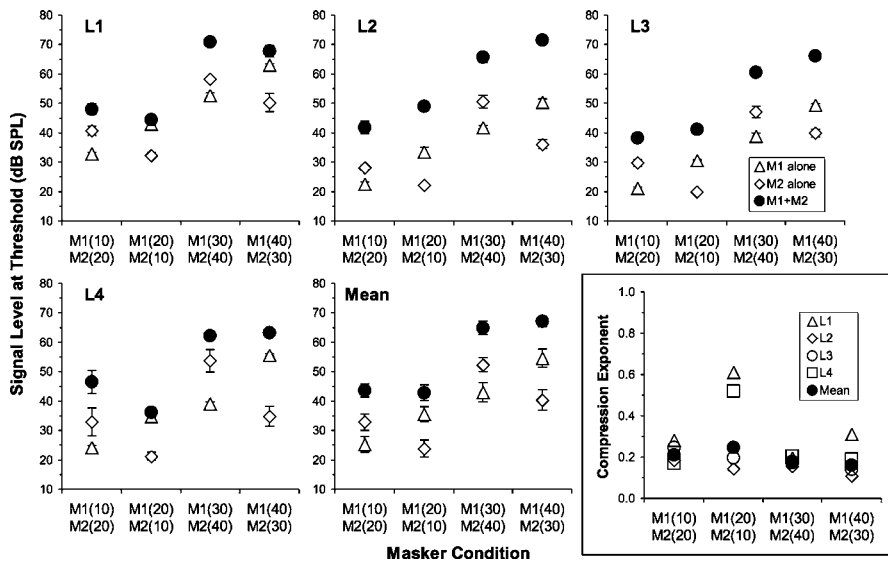
FIG. 1. The results of Phase 2 of the experiment, showing signal thresholds in the presence of the single (open symbols) and combined (closed symbols) maskers. Error bars show standard errors. The lower right panel shows compression exponents derived from the data (see the text for details).

masking was observed, resulting in an apparent difference between that condition and the M1(10)/M2(20) condition. This could be interpreted as evidence that the higher-level M1 was affecting the influence of the lower-level M2 for these two listeners. However, in the same subjects at the higher level, and in the other two listeners at both levels, no such evidence was found. Overall, as indicated by the statistical analysis, there was no evidence that the combined masker thresholds are affected by the order of presentation of the maskers.

B. Derivation of compression exponents

The technique for deriving compression exponents from the signal threshold data was the same as that used by Plack and O'Hanlon (2003). It was assumed that the ratio of the internal (i.e., postcochlear) signal magnitude to the internal (or effective) masker magnitude is a constant at signal threshold. It was also assumed that the internal signal magnitude is a power-law transformation of physical signal intensity. Hence,

$$I_M = kS_M^c, \tag{1}$$

where I_M represents the internal effect of the masker, S_M is the physical signal intensity at masked threshold, c is the compression exponent, and k is a constant. It was assumed further that the effect of combining two maskers is a linear summation of their individual effects. Hence,

$$I_{M1+M2} = I_{M1} + I_{M2}. \tag{2}$$

Substituting Eq. (1) in Eq. (2), and factoring out the constant k, leaves

$$S_{M1+M2}^c = S_{M1}^c + S_{M2}^c. \tag{3}$$

If S_{M1+M2} (the signal intensity at threshold for the combined maskers), S_{M1} (the signal threshold for M1 alone), and S_{M2} (the signal threshold for M2 alone) are all known, it is possible to determine the compression exponent c. This was achieved using the SOLVER algorithm in Microsoft Excel.

The derived compression exponents are shown in Fig. 1. The exponents for the mean data were derived from the mean

signal thresholds, rather than being simply the mean of the compression exponents across the four listeners. The values for the higher-level stimuli, 0.18 for M1(30)/M2(40) and 0.16 for M1(40)/M2(30), are close to the values reported in previous studies (e.g., Oxenham and Plack, 1997; Plack and O'Hanlon, 2003). Also as expected, there is a tendency for the values to increase at low levels, particularly for listeners L1 and L4. A repeated-measures ANOVA [overall level $(10/20, 30/40) \times$ masker sequence (low M1 / high M2, high M1 / low M2)] conducted on the compression exponents revealed a nearly significant effect of overall level [$F(1,3)=9.08$, $p=0.057$] no effect of masker sequence [$F(1,3)=1.28$, $p=0.34$] and no interaction [$F(1,3)=2.66$, $p=0.20$].

IV. CONCLUSIONS

Overall, the results are consistent with the hypothesis that the masking effects of nonoverlapping forward maskers combine in a linear manner (although this is only tested for masker and signal levels up to about 70 dB SPL in the present study). Furthermore, the compression exponents derived from the signal thresholds are consistent with previous psychophysical and physiological (Ruggero et al., 1997; Yates et al., 1990) measures of BM compression, suggesting that the main nonlinearity in forward masking is the result of cochlear processing. After quasi-instantaneous cochlear compression, the effects of combining nonoverlapping maskers can be well described by a time-invariant linear system.

[1]Phase 1 provided the masker levels required to the test the hypothesis using the combined masker thresholds from Phase 2 (Sec. III A). Although the second analysis in terms of compression exponents (Sec. III B) could be done using signal thresholds measured with arbitrary masker levels in

Phase 2, Phase 1 is necessary to at least approximately match the effectiveness of the maskers. If one masker is much more effective than the other it will dominate masking in the combined case, and the derivation of the exponent will be unreliable.

Jesteadt, W., Bacon, S. P., and Lehman, J. R. (**1982**). "Forward masking as a function of frequency, masker level, and signal delay," J. Acoust. Soc. Am. **71**, 950–962.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Moore, B. C. J., and Glasberg, B. R. (**1983**). "Growth of forward masking for sinusoidal and noise maskers as a function of signal delay: Implications for suppression in noise," J. Acoust. Soc. Am. **73**, 1249–1259.

Munson, W. A., and Gardner, M. B. (**1950**). "Loudness patterns—A new approach," J. Acoust. Soc. Am. **22**, 177–190.

Oxenham, A. J., and Moore, B. C. J. (**1995**). "Additivity of masking in normally hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **98**, 1921–1934.

Oxenham, A. J., and Plack, C. J. (**1997**). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **101**, 3666–3675.

Plack, C. J., and Drga, V. (**2003**). "Psychophysical evidence for auditory compression at low characteristic frequencies," J. Acoust. Soc. Am. **113**, 1574–1586.

Plack, C. J., and O'Hanlon, C. G. (**2003**). "Forward masking additivity and auditory compression at low and high frequencies," J. Assoc. Res. Otolaryngol. **4**, 405–415.

Plack, C. J., and Oxenham, A. J. (**1998**). "Basilar-membrane nonlinearity and the growth of forward masking," J. Acoust. Soc. Am. **103**, 1598–1608.

Plack, C. J., Oxenham, A. J., and Drga, V. (**2002**). "Linear and nonlinear processes in temporal masking," Acta. Acust. Acust. **88**, 348–358.

Plack, C. J., Oxenham, A. J., and Drga, V. (**2006**). "Masking by inaudible sounds and the linearity of temporal summation," J. Neurosci. **26**, 8767–8773.

Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **101**, 2151–2163.

Yates, G. K., Winter, I. M., and Robertson, D. (**1990**). "Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range," Hear. Res. **45**, 203–220.

Differential reductions in acoustic startle document the discrimination of speech sounds in rats (L)

Owen R. Floody[a)]
Department of Psychology, Bucknell University, Lewisburg, Pennsylvania 17837

Michael P. Kilgard
Neuroscience Program, School of Behavioral and Brain Sciences, P.O. Box 830688, GR 41,
University of Texas at Dallas, Richardson, Texas 75083-0688

The intensity of a noise-induced startle response can be reduced by the presentation of an otherwise neutral stimulus immediately before the noise ("prepulse inhibition" or PPI). This effect has been used to study the detection of gaps and other stimuli, but has been applied infrequently to complex stimuli or the ability to discriminate among multiple stimuli. To address both issues and explore the potential of PPI, rats were presented a series of 5 tasks, most contrasting a pair of speech sounds. One of these (the "standard" stimulus) occurred frequently but rarely preceded the startle stimulus. The second occurred infrequently (as an "oddball") and always preceded a noise. In each such task, startle responses were inhibited more by the oddball than by the standard stimulus, usually within the first test. This suggests that PPI can be adapted to studies of the discrimination of speech and other complex sounds, and that this method can provide useful information on subjects' ability to discriminate with greater ease and speed than other methods. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2770548]

I. INTRODUCTION

Studies of "reflex modification" use changes in the latency or magnitude of reflex responses to assess information processing. In one variant, stimuli that precede a loud noise can decrease the size of the subsequent startle response, a change described as "prepulse inhibition" (PPI). The present study was designed to determine if PPI can be adapted to study the processing of speech sounds by animals.

PPI has been studied for more than 40 years (Hoffman and Searle, 1965). Early studies surveyed the impact of parameters including the intensity of the initial stimulus, or "prepulse," the intensity of the startle stimulus, and the duration of the interstimulus interval (Hoffman and Ison, 1980). The results permitted the use of PPI to explore the mechanisms underlying detection, for instance by assessing the impact of damage to auditory cortex on the inhibition produced by a variety of prepulses (Bowen *et al.*, 2003).

Currently, PPI is used in at least two overlapping contexts. First, it is used to measure the capacity for "sensorimotor gating" (Braff and Geyer, 1990). The concern here is mechanisms that act early in information processing to focus that processing in an automatic or "preattentive" way. Second, PPI is used to study later "attentive" steps in the selection of information for processing. In each of these cases, recent research has used PPI to test for information processing deficits in a variety of populations (e.g., Hawk *et al.*, 2003) and to explore some of the mechanisms that underlie deficits in information processing (e.g., Schell *et al.*, 2000).

Though PPI has been used extensively to study detection, it has been used much less often to study the discrimination of multiple stimuli from the background and each other. Further, of discrimination studies using PPI, most have used an explicit attentional manipulation to stimulate discrimination (e.g., Hawk *et al.*, 2003). If one wanted to use PPI to study discrimination in animals, it would seem necessary to manipulate task structure instead, for instance by creating prepulses that relate to the startle stimulus differently. To our knowledge, the first study to use such an approach exposed rats to tone pairs that ascended or descended in frequency (Clark *et al.*, 2000a). One stimulus helped to define the acoustic background: It was presented very frequently, but rarely preceded a noise. The other stimulus, or "oddball," was designed to stand out from this background: It was presented infrequently and always preceded a noise. Across a wide range of stimulus parameters, startle responses were inhibited more by the oddball than the background stimulus. These and many subsequent results (Fitch and Peiffer, 2006) suggest that PPI can be adapted to the study of discrimination and represents a more flexible experimental tool than generally recognized.

We hoped to extend this success by refining PPI and applying it to speech sounds, stimuli that are more complex than those emphasized in past work. Speech sounds seem to merit attention both as examples of complex sounds and for any insights they can provide into mechanisms of speech perception. Already, animal models of speech perception seem to be making valuable contributions to our understanding of how acquired or developmental pathologies affect language processing (Fitch and Peiffer, 2006).

---
a)Electronic mail: ofloody@bucknell.edu

Oddball stimuli

Standard stimuli

Resulting tasks
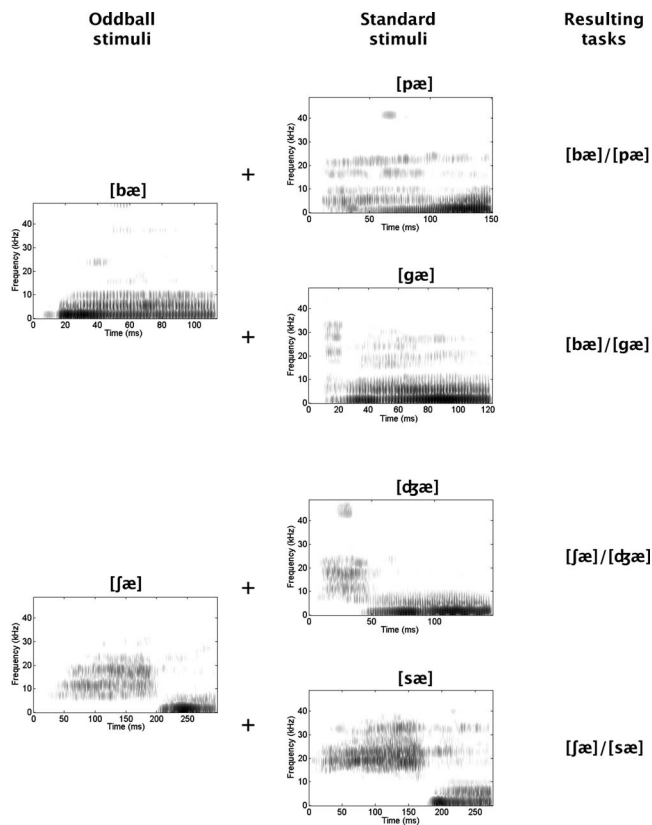
FIG. 1. Sound spectrograms of the 6 speech sounds used as stimuli. The 2 sounds that served as oddballs are described on the left. The 4 that served as standard stimuli are described in the middle. The combinations that defined the 4 discrimination tasks are specified on the right.

Past results show that animals can discriminate a variety of speech sounds (Brown and Sinnott, 2005). At the same time, these studies suggest that operant methods reveal such discriminations only with considerable time and effort (e.g., Kluender *et al.*, 1987). This obviously slows progress in research that uses speech perception in animals to study language processing. To address this issue, we studied speech perception in rats using an adaptation of PPI, a paradigm in which other discriminations can be acquired with impressive speed and that premininary results suggest can be applied to speech sounds (Clark *et al.*, 2000a,b).

II. METHODS

A. Animals and test chamber

The subjects were 12 female Sprague-Dawley rats averaging 139 days of age. Except during tests, each was housed individually in a wire-mesh cage with constant access to food and water. The colony was maintained at constant temperature and humidity, and on a reversed 12:12 h light:dark cycle. All methods and experimental treatments were approved by the University of Texas IACUC.

During testing, an animal was housed in a $20 \times 20 \times 20$ cm^3 wire-mesh cage in a $67 \times 67 \times 67$ cm^3 chamber lined with 5-cm acoustic foam. The cage was centered on a startle platform (Lafayette Instrument Co.) that uses a piezo-electric transducer to generate a continuous record of activity level. Sounds generated using an RP2.1 (Tucker-Davis Technologies) were delivered by a speaker (Optimus Bullet Horn Tweeter) mounted above the cage, about 20 cm from its center. Stimuli were adjusted for the speaker's frequency response using SigCal (Tucker-Davis Technologies). Sound intensities were measured using an ACO Pacific microphone (PS9200-7016) placed at a height approximating that of a standing rat's head.

B. Stimuli and startle response

Startle responses were elicited by 50-ms bursts of white noise at 102.0 dB. The waveform of each response (the peak to peak voltage within 500 ms of the noise) was sampled at 10 kHz using an RP2.1 and processed using MATLAB.

The stimuli also included six speech sounds that served as prepulses. These were derived from syllables (consonant-[æ]-[d]) spoken by a female native English speaker in a sound-proof chamber. During recording, these were sampled at 10 μs with 16-bit resolution. To adjust them to a rat's hearing, frequencies were doubled without changing the amplitude envelope (STRAIGHT vocoder, Kawahara *et al.*, 1998). To shorten prepulses and equate them for elements other than the initial consonant, all were truncated at 100 ms into the vowel. The intensity of the loudest 100 ms of each stimulus then was adjusted to about 60 dB (58.8–60.9).

The resulting stimuli included [bæ], [pæ], [gæ], [ʃæ], [dʒæ], and [sæ] (Fig. 1). Initial testing paired [bæ] with silence in a detection task. Later tasks required the discrimination of the sounds in the following pairs, presented in this order: [bæ]/[pæ], [bæ]/[gæ], [ʃæ]/[dʒæ], [ʃæ]/[sæ]. Order was not counterbalanced because of pilot data suggesting that the later tasks might be the more difficult, along with our initial uncertainty on rats' ability to master any of these tasks.

These pairings were designed to create tasks that could be solved by exploiting differences on particular dimensions (but see caveat below). Specifically, the first pair combines bilabial stop consonants differing in voicing ([b] voiced, [p] voiceless). These should be discriminable on the basis of voice onset time, a temporal difference. The second includes consonants differing in place of articulation, permitting a spectral discrimination revolving around the origins and trajectories of the second formant (low initially then ascending in [b], high initially then descending in [g]). The third contrasts a voiceless fricative with a voiced affricative. These should be discriminable using the duration of each consonant's initial noise burst (longer in [ʃ] than [dʒ]). This is a temporal difference, but of a different type from that distinguishing [bæ] and [pæ]. The fourth combines voiceless fricatives differing in place of articulation. These could be discriminated on the basis of the frequency range of the initial noise burst (relatively narrow and low for [ʃ], relatively broad and high for [s]), potentially representing a second type of spectral distinction. It is important to note, however, that the availability of a dimension does not guarantee the exploitation of that dimension. Natural speech sounds vary in multiple ways (Handel, 1993) and we cannot be sure how discriminations here were achieved.

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

O. R. Floody and M. P. Kilgard: Letters to the Editor   1885

C. Test procedures

Following 1 test on the detection task, each subject experienced 5 tests presenting [bæ] and [pæ], followed by 4 tests on each of the remaining 3 tasks. Each test included 3 phases. The first, in the detection task, consisted of 5 min of silence. In the other tasks, it involved the presentation at 1/s of a standard stimulus, which always was the second stimulus in a pair. Our goal here was to habituate any responses to this stimulus, causing it to recede into the background.

The second phase introduced the remaining speech sound (the first in each pair) and the startle stimulus. Ten of each were presented, with an interstimulus interval of 5 ms and spaced an average of 30 s (range=15–45 s) apart. Since the standard stimulus continued to appear at 1/s, it seems reasonable to refer to the newly introduced speech sound as an oddball (Clark *et al.*, 2000a). This phase was designed to highlight the relationship between the oddball and startle stimulus. The interstimulus interval was selected on the basis of pilot testing. It is shorter than generally optimal for PPI (e.g., Hoffman and Searle, 1965). This disparity may relate to the fact that much of the information available to support discriminations was concentrated prior to the vowel, and thus 105 ms or more before the startle stimulus.

Each test concluded with a 20-min "test phase" in which animals were exposed to a mixture of "cued" and "uncued" trials, each involving the presentation of the startle stimulus 5 ms after a prepulse. Based on prior usage (Clark *et al.*, 2000a), cued trials refer to those on which the oddball served as the prepulse. Uncued trials were those on which no stimulus (detection) or the standard stimulus (discriminations) preceded the startle stimulus. Each test included 10 blocks of 4 trials, each including 3 cued trials and one uncued, in random order and at intervals that again averaged 30 s. Except during these trials, "unreinforced" presentations of the standard stimulus continued at 1/s. Considering this, uncued trials are those on which there should be no prepulse that stands out from the background. Conversely, cued trials are those on which a prepulse should stand out, as long as the subject can distinguish the standard and oddball prepulses. Accordingly, one would expect greater PPI on cued than uncued trials, again as long as discrimination is possible.

D. Analysis

The critical data emerged from the test phase and included the average magnitudes of responses on cued and uncued trials. Average responses on the 1 day of detection testing were compared using a *t* test for dependent samples. The initial analysis of the discrimination data applied to each of these tasks an analysis of variance (ANOVA) using stimulus (standard, oddball) and day of testing (5 days for [bæ]/[pæ], 4 days for others) as within-subject factors. Ratios of the responses on cued and uncued trials (cued/uncued) were used to describe performance on each task and to compare levels of performance on the 4 discrimination tasks. Each data set was tested for heterogeneity of variance. Where found, significant heterogeneity was eliminated by logarithmic transformation prior to ANOVA. Throughout, a probability of 0.05 was used to define significance.
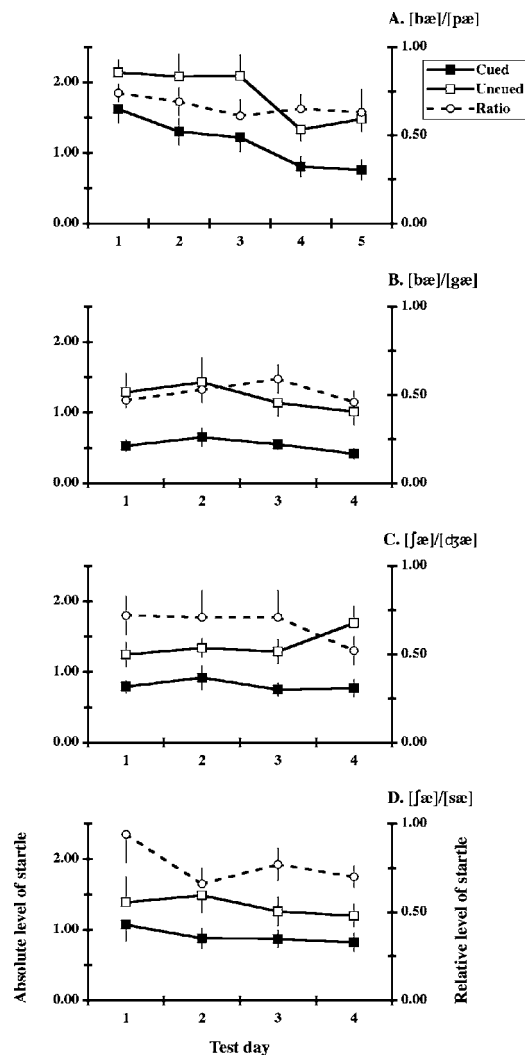
FIG. 2. Solid lines depict daily mean (±SEM) levels of startle (represented in volts on left axis) on cued and uncued trials (filled and open squares, respectively). Dashed lines depict mean cued/uncued ratios (scale on the right). Successive test days are represented on the horizontal axis. As indicated, panels (a)–(d) describe performance on the four tasks requiring the discrimination of speech sounds.

III. RESULTS

On the only day of detection testing, mean levels of startle on cued and uncued trials were 1.60 (SEM=0.28) and 3.52 (0.40), respectively. A *t* test for dependent samples confirmed a reliable reduction in the magnitude of startle on cued trials [$t(11)=8.25$, $p<0.001$, two-tailed].

The first discrimination task included 5 days of training using [bæ] and [pæ] as prepulses. ANOVA on these data revealed a reliable main effect of stimulus [$F(1,11)=26.19$, $p<0.001$], reflecting reduced levels of startle on cued trials [Fig. 2(a)]. This analysis also revealed a reliable main effect of days [$F(4,44)=10.20$, $p<0.001$]. Together with the absence of a reliable interaction, this reflects similar declines in cued and uncued responses over the 5 days of testing. These results are consistent with cued/uncued ratios that were stable and averaged significantly below 1 [mean=0.66, SEM=0.06, $t(11)=5.59$, $p<0.001$, one-sample *t* test, two-tailed].

O. R. Floody and M. P. Kilgard: Letters to the Editor

Analysis of the data from the remaining 3 tasks revealed similar results. In each, ANOVA revealed a reliable stimulus effect $[F(1,11) \geq 13.99, p \leq 0.003]$, reflecting lower levels of startle on cued trials [Figs. 2(b)–2(d)]. In no case was the main effect of day or the interaction significant. Again, these results are consistent with cued/uncued ratios that were stable over days, with averages consistently below 1 $[t(11) \geq 2.93, p \leq 0.014]$.

Though task order was not counterbalanced, we compared performance across the 4 discrimination tasks. This analysis subjected the cued/uncued ratios to a task×day ANOVA, including 4 days of testing on each task (the last 4 of those on [bæ]/[pæ]). This revealed a reliable main effect of task $[F(3,33)=4.04, p=0.015]$. This was clarified by the comparison of ratios averaged over days using the Tukey test. This revealed just one reliable difference, between the tasks with the highest and lowest average ratios ([ʃæ]/[sæ] and [bæ]/[gæ], respectively; $p < 0.05$, Fig. 2).

IV. DISCUSSION AND CONCLUSIONS

Our results are consistent with evidence documenting the value of PPI in studies of sensory function (e.g., Wecker et al., 1985). They also add to a more recent and smaller literature suggesting the utility of discriminative forms of PPI in studies of stimulus discrimination, sensorimotor gating, and attentional processes (e.g., Clark et al., 2000a; Hazlett et al., 2001). They extend these studies by demonstrating how easily PPI can be applied to the discrimination of complex sounds in animals. The paradigm described here elicited reliable detection within a single 30-min test. Further, it revealed reliable discriminations in each of 4 tasks presenting different pairs of speech sounds. As in the case of detection, discrimination on each of these was achieved rapidly, most or all within the first test.

The rapidity with which these discriminations were displayed is consistent with previous descriptions of PPI as reflexive and unlearned (e.g., Hoffman and Ison, 1980). On the other hand, some of our data are equivocal on this point, suggesting that stable discrimination on our most difficult task may not have been achieved until the second test day (data not shown). This suggests that perceptual or other learning may be required, at least when PPI is used to monitor some, possibly difficult, discriminations. This is consistent with the changes over tests and conditions seen in some previous studies (Crofton et al., 1990).

These results confirm the ability of nonhuman animals to discriminate human speech sounds. In itself, this finding is not novel (e.g., Brown and Sinnott, 2005). However, the ease with which discriminations were achieved here contrasts with the results of most past studies of speech processing by animals (e.g., Sinnott and Mosteller, 2001). Some of this contrast probably reflects our use of natural speech sounds, which incorporate more potential cues for discrimination than the synthetic stimuli used in much recent operant work (e.g., Sinnott and Mosteller, 2001). But even operant studies using natural speech sounds have required many trials for the emergence of reliable discriminations (e.g., Kluender et al., 1987). This suggests that stimulus features account for only

part of the difference in efficiency between operant methods and PPI. Consequently, our results support the application of PPI to tests of complex sound discrimination in animals, suggesting that at least some tasks could be studied more efficiently with some variant of reflex modification than with the operant methods that currently dominate the field.

Bowen, G. P., Lin, D., Taylor, M. K., and Ison, J. R. (**2003**). "Auditory cortex lesions in the rat impair both temporal acuity and noise increment thresholds, revealing a common neural substrate," Cereb. Cortex **13**, 815–822.

Braff, D. L., and Geyer, M. A. (**1990**). "Sensorimotor gating and schizophrenia," Arch. Gen. Psychiatry **47**, 181–188.

Brown, C. H., and Sinnott, J. M. (**2005**). "Cross-species comparisons of vocal perception," in *Listening to Speech: An Auditory Perspective*, edited by S. Greenberg and W. Ainsworth (Oxford U.P., New York), Chap. 12, pp. 183–201.

Clark, M. G., Rosen, G. D., Tallal, P., and Fitch, R. H. (**2000a**). "Impaired processing of complex auditory stimuli in rats with induced cerebrocortical microgyria: An animal model of developmental language disabilities," J. Cogn. Neurosci. **12**, 828–839.

Clark, M. G., Tallal, P., Rosen, G. D., Peiffer, A. M., and Fitch, R. H. (**2000b**). "Impaired perception of speech stimuli in rats with cerebrocortical microgyria," Abstr. Soc. Neurosci. **26**, 74.

Croften, K. M., Dean, K. F., Sheets, L. P., and Peele, D. B. (**1990**). "Evidence for an involvement of associative conditioning in reflex modification of the acoustic startle response with gaps in background noise," Psychobiol. **18**, 467–474.

Fitch, R. H., and Peiffer, A. M. (**2006**). "Behavioral consequences of focal anomalies in the cerebral cortex," in *The Dyslexic Brain: New Pathways in Neuroscience Discovery*, edited by G. D. Rosen (Erlbaum, Mahwah, NJ), pp. 259–288.

Handel, S. (**1993**). *Listening: An Introduction to the Perception of Auditory Events* (MIT, Cambridge, MA).

Hawk, L. W., Jr., Yartz, A. R., Pelham, W. E., Jr., and Lock, T. M. (**2003**). "The effects of methylphenidate on prepulse inhibition during attended and ignored prestimuli among boys with attention-deficit hyperactivity disorder," Psychopharmacol. **165**, 118–127.

Hazlett, E. A., Dawson, M. E., Schell, A. M., and Nuechterlein, K. H., (**2001**). "Attentional stages of information processing during a continuous performance test: A startle modification analysis," Psychophysiology **38**, 669–677.

Hoffman, H. S., and Ison, J. R., (**1980**). "Reflex modification in the domain of startle: I. Some empirical findings and their implications for how the nervous system processes sensory input," Psychol. Rev. **87**, 175–189.

Hoffman, H. S., and Searle, J. L. (**1965**). "Acoustic variables in the modification of startle reaction in the rat," J. Comp. Physiol. Psychol. **60**, 53–58.

Kawahara, H., de Cheveigne, A., and Patterson, R. D. (**1998**). "An instantaneous-frequency-based pitch extraction method for high-quality speech transformation: revised TEMPO in the STRAIGHT-suite," Proc. 5th Int. Conf. on Spoken Language Processing (ICSLP'96), Sydney.

Kluender, K. R., Diehl, R. L., and Killeen, P. R. (**1987**). "Japanese quail can learn phonetic categories," Science **237**, 1195–1197.

Schell, A. M., Wynn, J. K., Dawson, M. E., Sinaii, N., and Niebala, C. B. (**2000**). "Automatic and controlled attentional processes in startle eyeblink modification: Effects of habituation of the prepulse," Psychophysiology **37**, 409–417.

Sinnott, J. M., and Mosteller, K. W. (**2001**). "A comparative assessment of speech sound discrimination in the Mongolian gerbil," J. Acoust. Soc. Am. **110**, 1729–1732.

Wecker, J. R., Ison, J. R., and Foss, J. A. (**1985**). "Reflex modification as a test for sensory function," Neurobehav. Toxicol. Teratol. **7**, 733–738.

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

O. R. Floody and M. P. Kilgard: Letters to the Editor   1887

Rayleigh scattering of acoustic waves in rigid porous media

Claude Boutin[a]

*Laboratoire Geomateriaux, DGCB URA CNRS 1652, Ecole Nationale des Travaux Publics de l'Etat, 69518
Vaulx-en-Velin Cedex, France*

This paper describes the long wave scattering effect in gas saturated porous media using the homogenization method. To investigate the deviation from the continuum description, the multiscale asymptotic expansions are developed up to the third order. The leading (zeroth) order leads to the Biot-Allard continuum description. The correction of first order induces nonlocal terms in the dynamic Darcy law and thermal behavior, without modifying the wave characteristics. The correction of second order introduces additional dispersion effects on the velocity and attenuation. This theoretical approach is illustrated by analytical results in the simple case of a periodic array of slits. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2756755]

I. INTRODUCTION

This paper is devoted to the low frequency scattering of acoustic waves propagating in heterogeneous media made of air and motionless inclusions. This phenomenon occurs in noise-absorbing materials consisting of air-saturated porous media of sufficiently rigid and/or dense skeleton, and receives particular attention in this study. It may also appear in outdoor acoustics, if one considers, for instance, sound propagation through dense forest, crop fields, heavy rains, etc., or in ultrasonics applied to cellular or reticulated media (light concrete, foams, dried biological tissues, bones, etc.).

Usually, the equivalent fluid model (Zwikker and Kosten, 1949; Allard, 1993) efficiently describes the acoustics of such media, by means of the second compressional wave [P_2 wave in the sense of Biot (1956)]. This modeling captures the main physical effects, provided that the wavelength, λ, is significantly larger than the characteristic size, l, of the microstructure (for usual material l is about three times the largest pore diameter). However, when the wavelength is large, though not too large compared to the microstructure size, scattering phenomena begin to occur (Stanke and Kino, 1984). This long wave—or Rayleigh—scattering modifies the wave propagation and must be taken into account to improve the description of the effective properties in the corresponding frequency range. This idea is supported by experiments, e.g., Leclaire *et al.* (1996), showing that some deviations from modeling may result from scattering. The Rayleigh frequency range is bounded by the diffraction frequency, i.e., $\omega < \omega_d$, where ω_d is such that $\lambda(\omega_d) = 2\pi l$.

Rayleigh scattering in weakly dissipative heterogeneous media has been extensively studied and reviews of the literature on this topic can be found in Bond (1989), Ishimaru *et al.* (1997) and Sheng (1995). Nevertheless, only a few studies [for instance, Tournat *et al.* (2004)] focus on P_2 wave scattering. In fact, the specific features of P_2 waves invalidate the results on long wave scattering already established for other materials for several reasons:

(1) The physics couples viscous, thermal, and inertial effects in the transient regime at the local scale that are not accounted for in classical approaches.
(2) The strong dispersion of the P_2 wave—evolving from a diffusion wave, at low frequency, to a propagation wave at high frequency—induces significant differences compared with purely elastic or diffusive cases.
(3) The high contrast of properties between gas and skeleton makes the simplifying assumptions of quasihomogeneous material irrelevant (e.g., Hirsekorn, 1988).
(4) Finally, considering the usual porosity values, the Born approximation, assuming weak heterogeneity concentration (e.g., Gubernatis *et al.*, 1977), cannot be used.

Among several possible approaches, e.g., multiple scattering techniques (Kafesaki and Economou, 1999) Bloch waves (Turbe, 1982), etc., the method of homogenization of periodic media (Benssoussan *et al.*, 1978; Sanchez-Palencia, 1980; Bakhvalov and Panasenko, 1989), will be used in this paper. This multiscale asymptotic method enables one to determine the macroscopic description from knowledge of the physics at microscopic level, provided that a scale separation between macro- and microscale is fulfilled (Auriault, 1991). The procedure is generally restrained to the derivation of the first significant term that defines the equivalent continuum behavior. However, in the case of poor scale separation, the macroscopic continuum description has to be enriched by the next order terms that induce nonlocal effects [see Gambin and Kröner (1989) or Boutin (1996) for elastostatic cases]. For mechanical waves in elastic composite materials (Boutin and Auriault, 1993) it was shown that the so-derived correctors properly describe the Rayleigh scattering effects, that is, (i) a correction of polarization of the first order (whose amplitude linearly increases with the frequency), (ii) a dispersion of velocity of the second order (increasing with the square of the frequency), and (iii) an apparent attenuation of the third order, varying according to the cube of the frequency. A similar approach applied to thermal waves (Boutin, 1995) shows that the complex corrector of the effective diffusivity induces perturbations of velocity and attenuation,

---

[a]Electronic mail: claude.boutin@entpe.fr

which both increase linearly with frequency. Recently, the static Darcy's law correctors have been established by Auriault *et al.* (2005) and the adiabatic acoustic regime was investigated in Boutin and Bazaille (2005).

The paper is organized as follows. Section II is devoted to the principle of the homogenization method and to the basic physical assumptions concerning the medium. Section III presents the macroscopic description up to third order, and focus on the physical meaning of the correctors. In Sec. IV, the perturbations of plane wave are examined. Finally, a simple analytical example is given.

II. HOMOGENIZATION APPLIED TO THE ACOUSTICS OF GAS SATURATED POROUS MEDIA

A. Homogenization principle

The macroscopic representation of heterogeneous media makes sense only if there is a scale separation. This implies (Auriault, 1991) the following:

(1) The material is regular enough to show a representative volume element. This is expressed by considering that the material is composed of repeated identical cells of characteristic size l.
(2) The phenomenon must vary according to a size L larger than l. In acoustics, L is related to the wavelength by $L = \lambda/2\pi$ (Boutin and Auriault, 1990).

In order to catch the variations at the well distinct lengths L and l, two space variables are introduced: x for the macrovariations and y for the microvariations, x and y being related by the scale ratio $\epsilon = l/L \ll 1$; $y = \epsilon^{-1}x$. It is worth mentioning that, for a given medium, the actual physical scale ratio ϵ varies according to the wavelength, and therefore to the frequency.

Moreover, the small parameter ϵ suggests seeking variables (in the present case: the pressure p, the temperature θ, and the velocity v) in the form of asymptotic expansions in powers of ϵ:

$$p(x,y) = \sum_0^\infty \epsilon^i \, {}^ip(x,y), \quad \theta(x,y) = \sum_0^\infty \epsilon^i \, {}^i\theta(x,y), \qquad (1)$$

$$v(x,y) = \sum_0^\infty \epsilon^i \, {}^iv(x,y),$$

where ϵ^i means ϵ to the power i, while ${}^ip(x,y)$ is the ith term of the expansion of $p(x,y)$, etc. The scale separation and material periodicity induce the same periodicity for the physical quantities, thus the terms ${}^ip$, ${}^i\theta$, and ${}^iv$ are Ω-periodic according to the variable y. The homogenization proceeds in three steps:

(1) Performing a physical analysis and rescaling the equations, using powers of ϵ for expressing the order of magnitude of the dimensionless terms (see Sec. III B).
(2) Two-scale expansions are introduced in the rescaled two-scale equations and the terms of same power in ϵ are identified (developed in Sec. II C).
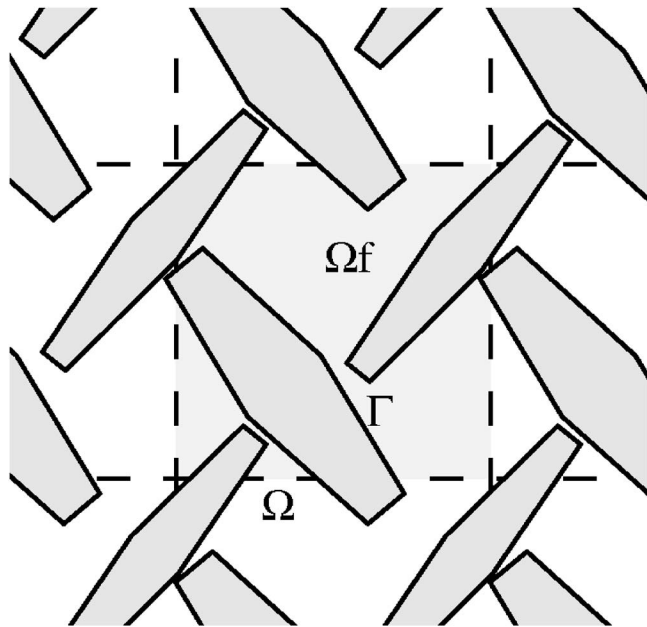
FIG. 1. Periodic cell Ω of porous media. Ω_f is the pore volume filled of gas. Γ is the air-solid interface. $\phi = \Omega_f/\Omega$ is the porosity.

(3) The resolution of the problems obtained in series (exposed Sec. III).

B. Harmonic gas motion porous media: Basic and rescaled equations

The gas saturating the motionless skeleton of porosity ϕ (ratio of the pore volume Ω_f to the volume Ω of the periodic cell),

$$\phi = \frac{\Omega_f}{\Omega},$$

is submitted to small harmonic perturbations (of frequency $f = \omega/2\pi$) from its equilibrium state (where the pressure, temperature, and density take the values P^e, T^e, and ρ^e). The variables describing the perturbations are the variations of pressure, p, temperature, θ, density, ρ, and the gas velocity v [$D(v)$ is the strain rate]. The parameters governing the motion and the heat transfer are the gas viscosity, μ, the thermal conductivity, κ, the specific heat ratio, γ, and the heat mass capacity c_p. Considering air as a perfect gas, one has the relation $c_p(1-1/\gamma) = P^e/T^e\rho^e$.

The linearized equations governing harmonic oscillations are given by the following (here and in the following the term $\exp(i\omega t)$ is omitted; ∇ stands for the gradient, Δ for the Laplacian, dot stands for contraction, double dots for double contraction, etc.):

(1) In the pores (Ω_f) of the periodic cell (Fig. 1):
Gas compressibility (mass balance):

$$\mathrm{div}(v) + i\omega\frac{\rho}{\rho^e} = 0. \qquad (2)$$

Navier-Stokes equation (momentum balance):

$$\mathrm{div}(2\mu D(v)) - \nabla p - i\omega\rho^e v = 0. \qquad (3)$$

Claude Boutin: Rayleigh scattering of poroacoustic waves   1889

Fourier equation (energy balance):

$$\text{div}(\kappa \nabla \theta) - i\omega(\rho^e c_p \theta - p) = 0. \tag{4}$$

State equation of the gas

$$\frac{p}{P^e} = \frac{\rho}{\rho^e} + \frac{\theta}{T^e}. \tag{5}$$

(2) On the gas–solid interface Γ, the boundary conditions are (here and in the following the notation $_{/\Gamma}$ means the value on Γ):
Adherence condition of the gas:

$$v_{/\Gamma} = 0. \tag{6}$$

Isothermal condition imposed by the weak thermal impedance of air, $\sqrt{\kappa \rho^e c_p}$, compared to that of the solid:

$$\theta_{/\Gamma} = 0. \tag{7}$$

During wave propagation, the pressure and the density oscillate according to the wavelength whereas the velocity and temperature vary at the pore scale. Then, using the wavelength as reference length, the Navier-Stokes and Fourier equations (3) and (4) have to be rescaled, the mass balance (2) being unchanged. The richest case is obtained when the pore size is of the same order as the viscous and thermal skin depths, i.e., $|\delta_v| = O(l)$ where $\delta_v = \sqrt{\mu/i\omega\rho^e}$ and $|\delta_t| = O(l)$ where $\delta_t = \sqrt{\kappa/i\omega\rho^e c_p} = O(l)$ (this is consistent with the fact that viscous and thermal layer thicknesses are of the same order for air). This situation leads one to replace Eqs. (3) and (4) by the two following rescaled equations [see, for instance, Boutin *et al.* (1998)]:

$$\epsilon^2 \text{div}(2\mu D(v)) - \nabla p - i\omega\rho^e v = 0, \tag{8}$$

$$\epsilon^2 \text{div}(\kappa \nabla \theta) - i\omega(\rho^e c_p \theta - p) = 0. \tag{9}$$

Note that the gas state equation enables the elimination of the density, and that the divergence of shear stresses, combined with the mass balance, may be expressed as follows:

$$\text{div}(2\mu D(v)) = \mu[\Delta(v) + \nabla \text{div}(v)]$$
$$= \mu\left[\Delta(v) - i\omega\nabla\left(\frac{p}{P^e} - \frac{\theta}{T^e}\right)\right].$$

Thus, the system is driven by the three differential operators acting on the variables p, v, and θ:

$$G(p,v,\theta) = 0, \quad G(p,v,\theta) = \text{div}(v) + i\omega\left(\frac{p}{P^e} - \frac{\theta}{T^e}\right), \tag{10}$$

$$N(p,v,\theta) = 0, \quad N(p,v,\theta) = -\nabla p - i\rho^e \omega v$$
$$+ \epsilon^2\mu\left[\Delta(v) - i\omega\nabla\left(\frac{p}{P^e} - \frac{\theta}{T^e}\right)\right], \tag{11}$$

$$F(p,\theta) = 0, \quad F(p,\theta) = i\omega(p - \rho^e c_p \theta) + \epsilon^2 \text{div}(\kappa \nabla \theta). \tag{12}$$

C. Homogenization process

Using the two space variables x and $y = \epsilon^{-1}x$, the gradient ∇ is changed into $\epsilon^{-1}\nabla_y + \nabla_x$, and the Laplacian Δ becomes $\epsilon^{-2}\Delta_y + 2\epsilon^{-1}\Delta_{yx} + \Delta_x$, with $\Delta_{yx} = \partial^2_{x_i y_i}$. Consequently, the operators G, N, F expressed with the two space variables are changed into G_{xy}, N_{xy}, F_{xy}, whose expansions are given by Eqs. (13)–(15):

$$\begin{cases} G_{xy}(p,v,\theta) = \epsilon^{-1}G^{-1}(v) + G^0(p,v,\theta), \\ G^{-1}(v) = \text{div}_y(v), \\ G^0(p,v,\theta) = \text{div}_x(v) + i\omega\left(\dfrac{p}{P^e} - \dfrac{\theta}{T^e}\right), \end{cases} \tag{13}$$

$$\begin{cases} N_{xy}(p,v,\theta) = \epsilon^{-1}N^{-1}(p) + N^0(p,v) + \epsilon N^1(p,v,\theta), \\ \qquad\qquad + \epsilon^2 N^2(p,v,\theta) \\ N^{-1}(p) = -\nabla_y p, \\ N^0(p,v) = -\nabla_x p - i\omega\dfrac{\rho^e}{\mu}\mu v + \Delta_y(\mu v), \\ N^1(p,v,\theta) = 2\Delta_{yx}(\mu v) - i\omega\dfrac{\mu}{P^e}\nabla_y\left(p - \dfrac{P^e}{T^e}\theta\right), \\ N^2(p,v,\theta) = \Delta_x(\mu v) - i\omega\dfrac{\mu}{P^e}\nabla_x\left(p - \dfrac{P^e}{T^e}\theta\right), \end{cases} \tag{14}$$

$$\begin{cases} F_{xy}(p,\theta) = F^0(p,\theta) + \epsilon F^1(p) + \epsilon^2 F^2(p), \\ F^0(p,\theta) = i\omega p - i\omega\rho^e c_p\theta + \kappa\Delta_y(\theta), \\ F^1(\theta) = 2\kappa\Delta_{yx}(\theta), \\ F^2(\theta) = \kappa\Delta_x(\theta). \end{cases} \tag{15}$$

Introduce now the expansions (1) of p, v, and θ in G_{xy}, N_{xy}, and F_{xy} and identify terms of identical power of ϵ. These balance equations become series in power of ϵ identically equal to zero whatever $\epsilon \ll 1$, so that each term must vanish. The so-derived equations at each order—combined with the adherence and isothermal conditions on Γ satisfied by each $^i v$ and $^i \theta$—lead to a series of problems to be solved recurrently. Section III presents the resolution up to the second order.

III. DESCRIPTION UP TO THE SECOND ORDER

A. Macroscopic mass balances

Before any resolution let us first mention that each mass balance of order $i \geq 0$, reads

$$\{G^{-1}(^{i+1}v) + G^0(^i p, {}^i v, {}^i \theta) = 0;$$
$$^{i+1}v_{/\Gamma} = 0; \quad ^{i+1}v \ \Omega\text{-periodic}\}.$$

The local periodicity introduces a compatibility condition obtained by integrating the local mass balance on the pore cell volume:

$$\int_{\Omega_f}\left\{\text{div}_y(^{i+1}v) + \text{div}_x(^i v) + i\omega\left(\frac{^i p}{P^e} - \frac{^i \theta}{T^e}\right)\right\}d\Omega = 0.$$

From the divergence theorem, the periodicity and the adherence condition on Γ, the first term vanishes. Then, inverting y integration and x derivative,

$$\operatorname{div}_x\left(\int_{\Omega_f} {}^i v\, d\Omega\right) + i\omega\left(\int_{\Omega_f} \frac{{}^i p}{P^e}\, d\Omega - \int_{\Omega_f} \frac{{}^i \theta}{T^e}\, d\Omega\right) = 0.$$

Introducing the physical (i.e., observable) macrovariables,

$${}^i V = \frac{1}{\Omega} \int_{\Omega_f} \epsilon^i\, {}^i v\, d\Omega = \frac{\phi}{\Omega_f} \int_{\Omega_f} \epsilon^i\, {}^i v\, d\Omega,$$

$${}^i P = \frac{1}{\Omega_f} \int_{\Omega_f} \epsilon^i\, {}^i p\, d\Omega, \quad {}^i T = \frac{1}{\Omega_f} \int_{\Omega_f} \epsilon^i\, {}^i \theta\, d\Omega,$$

one obtains the macroscopic mass balance at order i,

$$\operatorname{div}_x({}^i V) + i\omega\phi\left(\frac{{}^i P}{P^e} - \frac{{}^i T}{T^e}\right) = 0. \tag{16}$$

B. Expressions of local fields

To go further we need to determine the local fields $({}^i p, {}^i v, {}^i \theta)$ by solving successively the harmonic viscothermal linear problems on the cell. At each order, the problem may be split into a leading set of equations related to the gas flow, and a set of equations related to the heat transfer induced by the gas pressure. The general scheme of the procedure and the first steps of the resolution are presented hereafter. The more technical steps are reported in Appendix A.

1. Pressure and temperature at the leading (zero) order

The momentum balance at order ϵ^{-1} reduces to $N^{-1}({}^0 p) = -\nabla_y\, {}^0 p = 0$, which gives

$${}^0 p(x,y) = {}^0 P(x).$$

This result introduced in the heat transfer at order ϵ^0, namely $F^0({}^0 P, {}^0 \theta) = 0$, leads to the differential system $\{S_t^0\}$ set on Ω_f, governing ${}^0 \theta$:

$$S_t^0 \begin{cases} i\omega\, {}^0 P(x) - i\omega\rho^e c_p\, {}^0 \theta + \kappa\Delta_y({}^0 \theta) = 0, \\ {}^0 \theta_{/\Gamma} = 0, \quad {}^0 \theta\ \Omega\text{-periodic.} \end{cases}$$

The solution of this linear problem, with ${}^0 P(x)$ as forcing term, may be expressed in the form:

$$\frac{{}^0 \theta(x,y)}{T^e} = {}^0 \pi\left(\frac{y}{\delta_t}\right)\frac{{}^0 P(x)}{P^e}. \tag{17}$$

The temperature distribution ${}^0 \pi$ [solution for ${}^0 P(x) = P^e/T^e$] is complex and depends on the local variable and the frequency through the dimensionless variable y/δ_t. In the isothermal regime, reached at low frequency ($y/\delta_t \to 0$), the temperature vanishes and ${}^0 \pi(y/\delta_t) \to 0$. In the adiabatic regime, reached at high frequency ($y/\delta_t \to \infty$), the temperature tends to the uniform value $(1 - 1/\gamma)$.

2. Local velocity, pressure, and temperature at the first order

The local velocity, ${}^0 v$, and the pressure of the first order, ${}^1 p$, are obtained from the following system of equations $\{S_v^0\}$ set on Ω_f:

$$\{N^{-1}({}^1 p) + N^0({}^0 P, {}^0 v) = 0, \quad G^{-1}({}^0 v) = 0,$$

$${}^0 v_{/\Gamma} = 0, \quad \text{and}\quad {}^0 v, {}^1 p\ \Omega\text{-periodic}\}$$

which reads explicitly:

$$S_v^0 \begin{cases} -\nabla_y\, {}^1 p - \nabla_x\, {}^0 P - i\omega\dfrac{\rho^e}{\mu}\mu^0 v + \Delta_y(\mu^0 v) = 0, \\ \operatorname{div}_y({}^0 v) = 0, \\ {}^0 v_{/\Gamma} = 0, \quad {}^0 v,\ \text{and}\ {}^1 p\ \Omega\text{-periodic.} \end{cases}$$

This set defines the linear dynamic permeability problem with $\nabla_x\, {}^0 P$ as forcing term, whose solution is as follows [see Auriault (1980)]:

$$\mu^0 v(x,y) = {}^0 k\left(\frac{y}{\delta_v}\right)\nabla_x\, {}^0 P,$$
$$\tag{18}$$
$${}^1 p(x,y) = {}^1 \alpha\left(\frac{y}{\delta_v}\right)\nabla_x\, {}^0 P + \widehat{{}^1 p}(x).$$

The tensors ${}^0 k$ and ${}^1 \alpha$ are constituted by the three velocity and pressure distributions $(-{}^0 k^i/\mu, {}^1 \alpha^i)$ corresponding to unit pressure gradient in the three directions, $\nabla_x\, {}^0 P(x) = e_i$. The solutions $({}^0 k^i, {}^1 \alpha^i)$ are complex and depend on the dimensionless variable y/δ_v. In the quasistatic (low frequency) regime ($l/\delta_v \to 0$), the inertial effect vanishes and ${}^0 k$ tends to be real. At high frequency ($l/\delta_v \to \infty$), the inertia dominates and ${}^0 k$ tends to a purely complex tensor. Note that the pressure is defined up to a constant value $\widehat{{}^1 p}$ and that, by construction, ${}^1 \alpha$ is of zero mean value on the cell.

Now, the first corrector ${}^1 \theta$ of the temperature is derived from the heat transfer at order ϵ governed by the differential system $F^0({}^1 p, {}^1 \theta) + F^1({}^0 \theta) = 0$. This leads to the system $\{S_t^1\}$ set on Ω_f:

$$S_t^1 \begin{cases} i\omega\, {}^1 p(x,y) - i\omega\rho^e c_p\, {}^1 \theta + \kappa\Delta_y({}^1 \theta) + 2\kappa\Delta_{yx}({}^0 \theta) = 0, \\ {}^1 \theta_{/\Gamma} = 0, \quad {}^1 \theta\ \Omega\text{-periodic.} \end{cases}$$

The previously determined fields ${}^0 \theta$ and ${}^1 p$ [Eqs. (17) and (18)] introduce two independent forcing terms:

(1) $\widehat{{}^1 p}(x)$, inducing an identical problem than $\{S_t^0\}$ except that $\widehat{{}^1 p}(x)$ replaces ${}^0 P(x)$ and
(2) $\nabla_x\, {}^0 P$, linked with ${}^1 \alpha$ and $\nabla_y\, {}^0 \pi$ [coming from the equality: $\Delta_{yx}({}^0 \theta) = (T^e/P^e)\nabla_y^0 \pi \cdot \nabla_x\, {}^0 P$].

Therefore, by linearity, the solution is in the form:

$$\frac{{}^1 \theta(x,y)}{T^e} = {}^0 \pi\left(\frac{y}{\delta_t}\right)\frac{\widehat{{}^1 p}(x)}{P^e} + {}^1 \pi\left(\frac{y}{\delta_t}, \frac{y}{\delta_v}\right) \cdot \frac{\nabla_x\, {}^0 P}{P^e}. \tag{19}$$

The tensor of temperature distribution ${}^1 \pi$ is constituted by the three solutions ${}^1 \pi^i$ corresponding to pressure gradient in the three directions, $\nabla_x\, {}^0 P(x) = (P^e/T^e)e_i$. These particular solutions are complex and depend on the local variable and frequency through the two dimensionless variables y/δ_t and y/δ_v. This corresponds to a nonlocal transient equilibrium with thermoviscous coupling.

3. Local fields at the two next orders

Following the process, the pressure and velocity, ^2p and ^1v, are determined from the set $\{S_v^1\}$

$$S_v^1 \begin{cases} N^{-1}(^2p) + N^0(^1p,\,^1v) + N^1(^0P,\,^0v) = 0, \\ G^{-1}(^1v) + G^0(^0P,\,^0v) = 0, \\ ^1v_{/\Gamma} = 0, \; ^1v, \text{ and } ^2p \; \Omega\text{-periodic.} \end{cases}$$

and the temperature $^2\theta$ is solution of $\{S_t^2\}$:

$$S_t^2 \begin{cases} F^0(^2p,\,^2\theta) + F^1(^1\theta) + F^0(^0\theta) = 0, \\ ^2\theta_{/\Gamma} = 0, \quad ^2\theta \; \Omega\text{-periodic.} \end{cases}$$

Finally, ^3p and ^2v are derived from $\{S_v^2\}$:

$$S_v^2 \begin{cases} N^{-1}(^3p) + N^0(^2p,\,^2v) + N^1(^1p,\,^1v) + N^2(^0P,\,^0v) = 0, \\ G^{-1}(^2v) + G^0(^1p,\,^1v) = 0, \\ ^2v_{/\Gamma} = 0, \quad ^2v, \quad \text{and} \quad ^1p \; \Omega\text{-periodic.} \end{cases}$$

The resolution of $\{S_v^1\}$, $\{S_t^2\}$, and $\{S_v^2\}$ is reported in Appendix A. Let us underline that, at each order, the same differential set as the previous order is recovered, except for new additive terms, involving the gradient of the solutions of lower order. Thus, once $(^{j+1}p,\,^jv,\,^j\theta)$, $0 \leq j < i$, have been determined, they become forcing terms in the problems related to $(^{i+1}p,\,^iv,\,^i\theta)$, and so on.

4. Local fields up to the second order

To sum up, the pressure, velocity, and temperature, read, up to the second order:

$$p(x,y) = {}^0P(x) + \epsilon[^{1}\widehat{p}(x) + {}^1\alpha\nabla_x{}^0P(x)]$$

$$+ \epsilon^2\left[{}^2\widehat{p}(x) + {}^1\alpha\,.\,\nabla_x{}^1\widehat{p}(x) + {}^2\alpha\,.\,.\,\nabla_x\nabla_x{}^0P(x)\right.$$

$$\left. + \frac{i\omega\mu}{P^e}{}^2\nu^0P(x)\right] + \cdots,$$

$$-\mu v(x,y) = {}^0k\,.\,\nabla_x{}^0P(x) + \epsilon\left[{}^0k\,.\,\nabla_x{}^1\widehat{p}(x)\right.$$

$$\left. + {}^1k\,.\,.\,\nabla_x\nabla_x{}^0P(x) + \frac{i\omega\mu}{P^e}{}^1n^0P(x)\right]$$

$$+ \epsilon^2\left[{}^0k\,.\,\nabla_x{}^2\widehat{p}(x) + {}^1k\,.\,.\,\nabla_x\nabla_x{}^1\widehat{p}(x)\right.$$

$$\left. + {}^2k\,\cdots\,\nabla_x\nabla_x\nabla_x{}^0P(x) + \frac{i\omega\mu}{P^e}{}^2n\,.\,\nabla_x{}^0P(x)\right]$$

$$+ \cdots,$$

$$(P^e/T^e)\,\theta(x,y) = {}^0\pi^0P(x) + \epsilon[{}^0\pi^1\widehat{p}(x) + {}^1\pi\,.\,\nabla_x{}^0P(x)]$$

$$+ \epsilon^2\left[{}^0\pi^2\widehat{p}(x) + {}^1\pi\,.\,\nabla_x{}^1\widehat{p}(x)\right.$$

$$\left. + {}^2\pi\,.\,.\,\nabla_x\nabla_x{}^0P(x) + \frac{i\omega\mu}{P^e}{}^2\zeta^0P(x)\right]$$

$$+ \cdots.$$

As previously, tensors $^{i+1}\alpha$ and ^ik, respectively of rank $i+1$ and $i+2$, are constituted by the 3^{i+1} particular solutions of the *purely* viscoinertial transient problems $\{S_v^i\}$ under the 3^{i+1} components of the forcing term $(\nabla_x)^{i+1}\,{}^0P$. All of them are complex and depend on the local variable and frequency through y/δ_v only (cf. Appendix A). More precisely,

(1) $(-^1k^{mn}/\mu,\,^2\alpha^{mn})$ are the 9 solutions of $\{S_v^1\}$ with $\nabla_x\nabla_x{}^0P(x) = e_m \otimes e_n$.
(2) $(-^2k^{pqr}/\mu,\,^3\alpha^{pqr})$ are the 27 solutions of $\{S_v^2\}$ with $\nabla_x\nabla_x\nabla_x{}^0P(x) = e_p \otimes e_q \otimes e_r$.

Tensors $^{i+1}\nu$ and ^in, of rank $i-1$ and i, are constituted by the 3^{i-1} particular solutions of the *coupled* thermoviscoinertial transient problems $\{S_v^i\}$ under the 3^{i-1} components of the forcing term $(\nabla_x)^{i-1}\,{}^0P$. Consequently, they are complex and depend on both y/δ_t and y/δ_v:

(1) $(-^1n/\mu,\,^2\nu)$ is the solution of $\{S_v^1\}$ with $(i\omega\mu/P^e)^0P(x) = 1$.
(2) $(-^2n^m/\mu,\,^3\nu^m)$ are the 3 solutions of $\{S_v^2\}$ with $(i\omega\mu/P^e)\nabla_x{}^0P(x) = e_m$.

Finally, tensors $^i\pi$ of rank i are constituted by the 3^i solutions of $\{S_t^i\}$ under the 3^i unit components of the forcing term $(T^e/P^e)(\nabla_x)^i\,{}^0P$. Except for $^0\pi$, solution of the purely thermoinertial transient problem $\{S_t^0\}$, $^i\pi$ for $i > 0$, involves thermoviscoinertial coupling, so that they are complex and depend on both y/δ_t and y/δ_v. This is also the case of the scalar $^2\zeta$, solution of $\{S_t^2\}$ under $(i\omega\mu/P^e)^0P(x) = 1$.

It is important to keep in mind that the fields observable in the reality are $^ip\epsilon^i$, $^iv\epsilon^i$, and $^i\theta\epsilon^i$. Thus, the actual physical tensors—independent of the scale ratio !—are $^ik\epsilon^i$, $^i\pi\epsilon^i$, $^i\alpha\epsilon^i$, and $^2n\epsilon^2$.

C. Continuum description and correctors

Knowledge of the local fields enables to express the macrovariables appearing in the macromass balances (16). The averaging of the local fields (of nonzero mean value) on the cell enables definition of the macroscopic tensors whose dimension and magnitude are, respectively, $^iK \sim O(^0K)l^i$, $^i\Pi \sim l^i$, $^2N \sim l^2$, and $^2Z \sim O(1)$. They can be determined (numerically) as soon as the physical parameters of the gas, the frequency, and the pore geometry are given. Note that the tensors issued from viscothermal coupled problems depend on both dimensionless frequencies ω/ω_c and ω/ω_t [see Eqs. (21) and (22) for the definition of ω_c and ω_t],

$$^iK\left(\frac{\omega}{\omega_c}\right) = \frac{1}{\Omega}\int_{\Omega_f}{}^ik\epsilon^i d\Omega, \quad ^0\Pi\left(\frac{\omega}{\omega_t}\right) = \frac{1}{\Omega_f}\int_{\Omega_f}{}^0\pi\epsilon^i d\Omega$$

and for $i > 1$,

$$^iN\left(\frac{\omega}{\omega_c},\frac{\omega}{\omega_t}\right) = \frac{1}{\Omega}\int_{\Omega_f}{}^in\epsilon^2 d\Omega,$$

$$^i\Pi\left(\frac{\omega}{\omega_t},\frac{\omega}{\omega_c}\right) = \frac{1}{\Omega_f}\int_{\Omega_f}{}^i\pi\epsilon^i d\Omega,$$

$$^2Z\left(\frac{\omega}{\omega_t},\frac{\omega}{\omega_c}\right) = \frac{1}{\Omega_f}\int_{\Omega_f}{}^2\zeta\epsilon^2 d\Omega.$$

Their main properties, especially symmetries, and relations enabling one to deduce *effective tensors* of the *second* order from *local solutions* of the *first* order, are established in Appendix B.

1. Biot-Allard modeling

The leading equations (i.e., zero order) are in agreement with the phenomenological approaches of Zwikker and Kosten (1949), Attenborough (1983), and Allard (1993). The equivalent continuum is described by the dynamic permeability tensor ${}^0K/\mu$ and the effective compressibility $\phi[1-{}^0\Pi]/P^e$ (in the following, to save notations, the index x is omitted for the macroscopic derivatives):

$$\begin{cases} \mathrm{div}({}^0V) + i\omega\phi\left[\dfrac{{}^0P}{P^e} - \dfrac{{}^0T}{T^e}\right] = 0, \\ \mu\,{}^0V = -{}^0K\cdot\nabla{}^0P, \\ \dfrac{{}^0T}{T^e} = {}^0\Pi\dfrac{{}^0P}{P^e}. \end{cases} \qquad (20)$$

It is necessary for the following to recall the features of the dynamic permeability (Auriault *et al.*, 1985). At low frequencies, viscous effects dominate and ${}^0K$ tends toward the real-valued intrinsic permeability, $\mathcal{K}$. At high frequencies, inertia dominates and ${}^0K$ tends towards a pure imaginary value, $\phi\mu/i\omega\rho^e\alpha_\infty$, where α_∞ is the tortuosity. Low and high frequency domains are delimited by a critical frequency derived by equalizing viscous and inertial effects of the macroscopic flow:

$$\omega_c = \frac{\phi\mu}{\mathcal{K}\rho^e\alpha_\infty}. \qquad (21)$$

As for the effective compressibility, at low frequencies the temperature variation vanishes, so that ${}^0\Pi \to 0$ and the effective compressibility tends toward the isothermal compressibility ϕ/P^e. At high frequencies, conduction effects are negligible, except in close proximity to the solid. The perturbations become quasiadiabatic so that ${}^0\Pi \to 1-1/\gamma$, and the compressibility tends toward the adiabatic value $\phi/\gamma P^e$. The thermal characteristic pulsation delimiting both regimes is related to the length Λ_t, defined—up to a geometric factor of the order of one—as the ratio of the volume to the surface of pore Ω_f/Γ (Champoux and Allard (1991)):

$$\omega_t = \frac{\kappa}{\Lambda_t^2\rho^e c_p}. \qquad (22)$$

Whereas viscous and thermal layer thicknesses are of the same order in air, the frequencies ω_c and ω_t can be rather different because the permeability essentially depends on the small ducts, while thermal transfer involves the whole pore volume. In consequence, $\omega_c > \omega_t$. Note also that the effective compressibility always contains a real part, and thus the change from isothermal to adiabatic conditions has consequences less drastic than the change from viscous to inertial regime.

2. Correctors to Biot-Allard modeling

The equations governing the two next orders provide the correctors to the continuum description.

First corrector:

$$\begin{cases} \mathrm{div}({}^1V) + i\omega\phi\left[\dfrac{{}^1P}{P^e} - \dfrac{{}^1T}{T^e}\right] = 0, \\ \mu\,{}^1V = -{}^0K\cdot\nabla{}^1P - {}^1K \mathbin{.\,.} \nabla\nabla{}^0P - \dfrac{i\omega\mu}{P^e}{}^1N\,{}^0P, \\ \dfrac{{}^1T}{T^e} = {}^0\Pi\dfrac{{}^1P}{P^e} + {}^1\Pi\cdot\nabla\dfrac{{}^0P}{P^e}. \end{cases} \qquad (23)$$

Second corrector:

$$\begin{cases} \mathrm{div}({}^2V) + i\omega\phi\left[\dfrac{{}^2P}{P^e} - \dfrac{{}^2T}{T^e}\right] = 0, \\ \mu\,{}^2V = -{}^0K\cdot\nabla{}^2P - {}^1K \mathbin{.\,.} \nabla\nabla{}^1P - {}^2K \ldots \nabla\nabla\nabla{}^0P \\ \qquad - \dfrac{i\omega\mu}{P^e}{}^2N\cdot\nabla{}^0P, \\ \dfrac{{}^2T}{T^e} = {}^0\Pi\dfrac{{}^2P}{P^e} + {}^1\Pi\cdot\nabla\dfrac{{}^1P}{P^e} + {}^2\Pi\mathbin{.\,.}\nabla\nabla\dfrac{{}^0P}{P^e} + \dfrac{i\omega\mu}{P^e}{}^2Z\dfrac{{}^0P}{P^e}. \end{cases}$$
$$(24)$$

These equations underline that a poor scale separation introduces deviations from the continuum description (strictly valid for infinite scale ratio) involving the successive gradients of pressure. The origin of those correctors lies in the terms neglected at preceding orders that become significant at the considered order. They bring unusual nonlocal terms and gas compressibility terms in both dynamic Darcy law and thermal behavior.

3. Isotropy or symmetric cell case: Cancellation of the first corrector of pressure, velocity, and temperature

In the case of macroscopic isotropy of the medium (up to the second order) or of cell presenting symmetry according to three orthogonal planes, tensors of odd rank, ${}^1K$ and ${}^1\Pi$ cancel out. Therefore, Eqs. (23) driving the first correctors reduces to the same as that of the leading order (20). Consequently, without loss of generality, ${}^1P, {}^1V, {}^1T$ can be canceled out, and the effective correction of the three variables is reported to the second order, that is:

$$P = {}^0P + {}^2P + \dots, \quad V = {}^0V + {}^2V + \dots, \quad T = {}^0T + {}^2T + \dots .$$

Furthermore, as any fourth rank isotropic tensor, ${}^2K$ takes the following form (a, b, c are three independent scalars and δ the Kronecker symbol):

$${}^2K_i^{jkl} = a\,\delta_i^j\delta_k^l + b\,\delta_i^k\delta_j^l + c\,\delta_i^l\delta_k^j$$

giving:

$${}^2K \ldots \nabla\nabla\nabla{}^0P = (a+b+c)\,\nabla(\Delta{}^0P).$$

Consequently, ${}^2K$ may be reduced to a scalar function ${}^2k = a+b+c$ and the velocity correctors read (${}^0k$ and ${}^2n$ are the

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Claude Boutin: Rayleigh scattering of poroacoustic waves   1893

scalar functions associated by the isotropic tensors ^0K and ^2N):

$$\mu^2 V = -{}^0k\, \nabla\, {}^2P - {}^2k\, \nabla\, (\Delta\, {}^0P) - \frac{i\omega\mu}{P^e}\,{}^2n\, \nabla\, {}^0P$$

$$= -{}^0k\, \nabla\, {}^2P + \frac{{}^2k}{{}^0k}\Delta(\mu^0V) - \frac{i\omega\mu}{P^e}\,{}^2n\, \nabla\, {}^0P.$$

After adding the terms of zero order, we have

$$\mu\left[V - \frac{{}^2k}{{}^0k}\Delta(V)\right] = -{}^0k\left[1 + \frac{i\omega\mu}{P^e}\frac{{}^2n}{{}^0k}\right]\nabla\, P + O(\epsilon^3). \quad (25)$$

The effective dynamic Darcy's law is enriched by two terms: (a) One linked to ^2k is of viscoinertial nature—it corresponds to a generalization of the Brinkmans law in the dynamic range. (b) The other linked to ^2n involves the thermoviscous coupling induced by the gas compressibility—it introduces a correction of permeability.

As for the thermal corrector in isotropic case, one obtains ($^2\pi$ is the scalar function associated to the isotropic tensor $^2\Pi$):

$$\frac{{}^2T}{T^e} = {}^0\Pi\frac{{}^2P}{P^e} + {}^2\pi\Delta\left(\frac{{}^0P}{P^e}\right) + \frac{i\omega\mu}{P^e}{}^2Z\frac{{}^0P}{P^e},$$

which gives, added to the zero-order terms:

$$\frac{T}{T^e} = \left[{}^0\Pi + \frac{i\omega\mu}{P^e}{}^2Z\right]\frac{P}{P^e} + {}^2\pi\Delta\left(\frac{P}{P^e}\right) + O(\epsilon^4).$$

This effective state equation for the gas includes nonlocal correction in the dynamic range with thermoviscous coupling effects linked with the compressibility.

4. General case: Cancellation of the first corrector of pressure

By combining the three equations of Eq. (20) governing the leading order, the velocity and the temperature may be eliminated, and the scalar wave equation expressed with the pressure ^0P only is derived:

$$-\mathrm{div}\left(\frac{{}^0K}{\mu}\,.\,\nabla^0P\right) + i\omega\frac{\phi[1 - {}^0\Pi]}{P^e}{}^0P = 0. \quad (26)$$

Similarly, the first order set [Eq. (23)] yields

$$-\mathrm{div}\left(\frac{{}^0K}{\mu}\,.\,\nabla^1P\right) + i\omega\frac{\phi[1 - {}^0\Pi]}{P^e}{}^1P = S^1({}^0P), \quad (27)$$

$$S^1({}^0P) = \mathrm{div}\left(\frac{{}^1K}{\mu}\,.\,.\,\nabla\nabla^0P\right) + \frac{i\omega}{P^e}{}^1N\,.\,\nabla^0P + i\omega\frac{\phi}{P^e}{}^1\Pi\,.\,\nabla^0P. \quad (28)$$

It is worth mentioning that, if the same differential operator (left-hand side) applies for ^0P and ^1P, the difference between zero- and first-order equations (26) and (27) comes from the source term on the right-hand side. In fact, Eq. (27) means that the first corrector field ^1P results from radiation of a density of source $S^1({}^0P)$ generated by the ^0P field. However, the first corrector presents generally a particular ener-

getic property. In composite media, it was shown—in static or dynamic elasticity or for thermal transfer—that the work of the first corrective term under the zero-order field is null (Boutin and Auriault, 1993; Boutin, 1995, 1996). This is shown below in the frame of poroacoustics. The source term $S^1({}^0P)$ reads

$$S^1({}^0P) = \frac{1}{\mu}{}^1K_j^{pq}{}^0P_{,pqj} + \frac{i\omega}{P^e}[{}^1N_p + \phi^1\Pi^p]{}^0P_{,p}.$$

However, the skew symmetry of ^1K and identity between ^1N and $\phi^1\Pi$ can be demonstrated, see Appendix B:

$$^1K_j^{pq} = -{}^1K_p^{jq} \quad \text{and} \quad {}^1N_p = -\phi^1\Pi^p. \quad (29)$$

Combining this identity with $^0P_{,pqj} = {}^0P_{,jqp}$, one deduces that

$$S^1({}^0P) = 0 \quad \text{then} \quad {}^1P = 0.$$

As a matter of fact, ^1P being governed by the same equations as ^0P, it can be canceled out without loss of generality. Nevertheless, contrary to the isotropic case, ^1V and ^1T do not necessarily disappear.

5. Effective correctors

To sum up, *in any case*, the governing equations [Eqs. (23) and (24)] can be replaced by the following.
First corrector:

$$^1P = 0, \quad \mu^1V = -{}^1K\,.\,\nabla\nabla^0P - \frac{i\omega}{P^e}{}^1N^0P, \quad (30)$$

$$\frac{{}^1T}{T^e} = {}^1\Pi\,.\,\nabla\frac{{}^0P}{P^e}.$$

Because of the identities (29), the macromass balance in Eq. (23) is necessarily satisfied.
Second corrector:

$$\mathrm{div}({}^2V) + i\omega\phi\left[\frac{{}^2P}{P^e} - \frac{{}^2T}{T^e}\right] = 0,$$

$$\mu^2V = -{}^0K\,.\,\nabla^2P - {}^2K\,...\,\nabla\nabla\nabla^0P - \frac{i\omega\mu}{P^e}{}^2N\,.\,\nabla^0P,$$

$$\frac{{}^2T}{T^e} = {}^0\Pi\frac{{}^2P}{P^e} + {}^2\Pi\,.\,.\,\nabla\nabla\frac{{}^0P}{P^e} + \frac{i\omega\mu}{P^e}{}^2Z\frac{{}^0P}{P^e}. \quad (31)$$

The equation governing the actual pressure corrector ^2P derived from Eq. (31) takes the form:

$$-\mathrm{div}\left(\frac{{}^0K}{\mu}\,.\,\nabla^2P\right) + i\omega\frac{\phi[1 - {}^0\Pi]}{P^e}{}^2P = S^2({}^0P) \quad (32)$$

with

$$S^2({}^0P) = \mathrm{div}\left(\frac{{}^2K}{\mu}\,...\,\nabla\nabla\nabla^0P\right) + \frac{i\omega}{P^e}\bigg[{}^2N\,.\,.\,\nabla\nabla^0P$$

$$+ \phi\left({}^2\Pi\,.\,.\,\nabla\nabla^0P + \frac{i\omega\mu}{P^e}{}^2Z^0P\right)\bigg]. \quad (33)$$

As shown previously, the term on the right-hand side acts as a density of source generated by the ^0P field. However, at

1894   J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Claude Boutin: Rayleigh scattering of poroacoustic waves

this order the source $S^2(^0P)$ does not cancel out and, in turn, a corrective ^2P field is radiated (according to the zero-order macrobehavior): We are back to the Rayleigh scattering, in which the passing of a long wave through heterogeneity generates sources that radiate perturbations. The interest of the homogenization is to replace these sources and their precise radiation by mean sources and equivalent fields at the considered order. Furthermore, theoretical expressions are provided for determining the effective coefficients from the knowledge of the microstructure.

6. Simplification in adiabatic (or isothermal) regime

The above presented description [Eqs. (20), (30), and (31)] may be simplified if the adiabatic regime is assumed in the gas. In that case, the conduction disappears and temperature, pressure, and density are related by

$$\frac{\theta}{T^e} = (1 - 1/\gamma)\frac{P}{P^e}, \quad \frac{\rho}{\rho^e} = \frac{P}{\gamma P^e}, \tag{34}$$

so that $^0\Pi = 1 - 1/\gamma$ and the tensors $^i\Pi$, ^2Z, ^1N disappear. Tensors ^0K, ^1K, and ^2K are the same as previously noted, since they are independent of the heat transfer. However, the tensor ^2N is modified by the absence of thermal coupling. The isothermal case, which implies $\theta = 0$, then $^i\Pi = 0$, and $\rho = P(\rho^e/P^e)$ leads to similar simplifications (with ^2N different than in the adiabatic case).

IV. RAYLEIGH SCATTERING OF PLANE WAVES

A. The continuum approximation

Consider a porous medium (isotropy is not needed) and investigate the scattering effect on an harmonic plane wave propagating in a given direction (of unit vector d). Denoting by $h_d = h_r + ih_i$ the complex wave number in this direction, the variations of the zero-order pressure take the form:

$$^0P = \widetilde{^0P} \exp(-ih_d x . d)\exp(i\omega t). \tag{35}$$

This plane wave is an eigenmode of the medium, and, from the zero-order balance Eq. (26),

$$(-ih_d)^2 = i\omega \frac{\phi[1 - {}^0\Pi]}{P^e}\frac{\mu}{K_d}$$

where

$$K_d = d_i {}^0K_i^j d_j.$$

The wave velocity C_d, wavelength λ_d, and attenuation ξ_d in direction d are related to the wave number by

$$h_d = \frac{\omega}{C_d}(1 - i\xi_d),$$

i.e.,

$$C_d = \frac{\omega}{h_r}, \quad \lambda_d = \frac{2\pi}{h_r}, \quad \xi_d = -\frac{h_i}{h_r}.$$

Because of the variations of the dynamic permeability with frequency recalled in Sec. III C 1, the P_2 wave shows a strong dispersion, evolving from

(1) a diffusion wave at low frequency ($\omega < \omega_c$):
$C_d \simeq \sqrt{(\omega/\omega_c)}\sqrt{(P^e/\alpha_\infty\rho^e)}$; $\xi_d = 1 - O(\sqrt{(\omega/\omega_c)})$.
(2) to a propagation wave at high frequency ($\omega > \omega_c$):
$C_d \simeq \sqrt{(\gamma P^e/\alpha_\infty\rho^e)}$; $\xi_d \sim \sqrt{(\omega_c/\omega)}$.

B. Scattering effect

As $^1P = 0$, the above-presented continuum description is valid up to a precision ϵ^2 for the pressure and ϵ for the flow and temperature. This is sufficient for long waves. For shorter waves the precision is improved by considering the higher order terms. With this aim, let us first calculate the source term $S^2(^0P)$. Replacing in Eq. (33) ^0P by its expression (35), gives

$$S^2(^0P) = \left\{\frac{1}{\mu}{}^2K_d(-ih_d)^4 + \frac{i\omega}{P^e}\left\{[{}^2N_d + {}^2\Pi_d\phi](-ih_d)^2\right.\right.$$
$$\left.\left. + \frac{i\omega\mu}{P^e}{}^2Z\phi\right\}\right\}\widetilde{^0P}\exp(-ih_d x . d)$$

where

$$^2K_d = d_i^2 K_i^{jkl} d_j d_k d_l, \quad {}^2N_d = d_i^2 N_i^j d_j, \quad {}^2\Pi_d = d_i^2 \Pi_i^j d_j.$$

This source is proportional (but not in phase) to ^0P. Thus, it acts as a forcing term loading the medium according to one of the eigenmodes, i.e., the plane wave in the direction d. This induces a *self-resonance* effect and subsequently the radiated field ^2P is *linearly amplified* as the wave progresses. Thus, the field ^2P satisfying Eq. (26) is in the form:

$$^2P = -Q_d(-ih_d)^2(-ih_d x . d)\widetilde{^0P}\exp(-ih_d x . d), \tag{36}$$

where the frequency-dependent complex coefficients, Q_d and $Q_d(-ih_d)^2$, read:

$$2Q_d = \frac{{}^2K_d}{K_d} + \frac{{}^2N_d + {}^2\Pi_d\phi}{\phi(1 - {}^0\Pi)} + {}^2Z\phi\frac{K_d}{[\phi(1 - {}^0\Pi)]^2} \tag{37}$$

and

$$2Q_d(-ih_d)^2 = \frac{i\omega\mu}{P^e}\left[\phi(1 - {}^0\Pi)\frac{{}^2K_d}{(K_d)^2} + \frac{{}^2N_d + {}^2\Pi_d\phi}{K_d}\right.$$
$$\left. + \frac{{}^2Z}{1 - {}^0\Pi}\right]. \tag{38}$$

Expression (36) means that the scattered field is coherent (consistently with the scale separation assumption) and radiates in the same direction as the zero-order field. Finally, up to the second order the macropressure reads:

$$^0P + {}^2P = [1 - Q_d(-ih_d)^2(-ih_d x . d)]\widetilde{^0P}\exp(-ih_d x . d)$$
$$= P(1 + O(\epsilon^3)).$$

At the same level of approximation (apparently within a distance x such as $O(|h_d|x.d) < 1/\sqrt{\epsilon}$, however this restriction can be removed by considering new boundary conditions beyond this distance), the macropressure can also be expressed as

$$P = \widetilde{^0P}(\exp(-i\widetilde{h_d}x . d) + O(\epsilon^3))$$

with

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Claude Boutin: Rayleigh scattering of poroacoustic waves   1895

$$\widetilde{h}_d = h_d[1 - Q_d(-ih_d)^2].$$

Thus the diffraction modifies the wave number h_d given by the continuum approximation. As expected, the correction is of the order of $Q_d(-ih_d)^2 = O(l^2/(\lambda/2\pi)^2) = O(\epsilon^2)$. This perturbation comes from the interference between the zero-order wave and the coherent and amplified wave generated by the induced source distribution. Writing Q_d on the polar form: $Q_d = |Q_d|\exp(i\chi)$, one derives the apparent velocity and attenuation factor by expressing explicitly that

$$\widetilde{h}_d = \frac{\omega}{\widetilde{C}_d}(1 - i\widetilde{\xi}_d) = \frac{\omega}{C_d}(1 - i\xi_d)\left[1 + |Q_d|(\cos\chi + i\sin\chi)\right.$$
$$\left. \times\left[\frac{\omega(1 - i\xi_d)}{C_d}\right]^2\right].$$

Algebra leads to

$$\widetilde{C}_d = C_d\left[1 - |Q_d|\left(\frac{\omega}{C_d}\right)^2[(1 - \xi_d^2)(\cos\chi + \xi_d\sin\chi)\right.$$
$$\left. + 2\xi_d(\sin\chi - \xi_d\cos\chi)]\right], \qquad (39)$$

$$\widetilde{\xi}_d = \xi_d + |Q_d|\left(\frac{\omega}{C_d}\right)^2[2\xi_d(1 + \xi_d^2)\cos\chi - (1 - \xi_d^4)\sin\chi].$$
$$\qquad (40)$$

Those expressions point out the difference between the scattering effect for poroacoustic waves and for elastic or thermal waves. Let us recall that in the two latter cases the analogue of Q_d is a purely real constant (i.e., $\chi = 0$) strictly related to the geometry of the microstructure [the reason is that—contrary to poroacoustics—in those cases, quasistatic conditions govern the local physics, see [Boutin and Auriault (1993), Boutin (1995), Chen and fish (2001)]. In elasticity, the zero-order velocity is a constant, and there is no attenuation (i.e., $\xi = 0$): The scattering leads to a velocity dispersion varying according to ω^2 and an apparent attenuation increasing according to ω^3. The thermal waves being dispersive ($C \sim \sqrt{\omega}$) and damped ($\xi = 1$), the scattering induces an additional velocity dispersion and attenuation both varying linearly with ω.

In poroacoustics the rules for the physical consequences of scattering are not so easy to draw, first, because of the change in the P_2 wave from a diffusion to a propagation mode, and, second, because Q_d is a frequency-dependent complex function. Consequently, the velocity and attenuation are both modified, but no general simple trends in the whole frequency range can be drawn for the frequency dependence.

In the viscous regime, the vanishing of the transient effects at the pore scale makes macroscopic tensors tend toward real value, then $Q_d \rightarrow Q_{d0}$ and $\chi \rightarrow 0$ and one obtains a correction identical to that of a diffusion wave:

$$\widetilde{C}_d \simeq C_d\left[1 + 2Q_{d0}\left(\frac{\omega}{C_d}\right)^2\right], \quad \widetilde{\xi}_d \simeq \xi_d + 4Q_{d0}\left(\frac{\omega}{C_d}\right)^2$$
$$\qquad (41)$$

with

$$C_d \simeq \sqrt{\frac{\omega}{\omega_c}}\sqrt{\frac{P^e}{\alpha_\infty\rho^e}}, \quad \xi_d = 1 - O\left(\sqrt{\frac{\omega}{\omega_c}}\right).$$

In the inertial regime (below diffraction), i.e., $\omega_c < \omega < \omega_d$, the wave propagates with damping and one may expect perturbation in both velocity and attenuation:

$$\widetilde{C}_d \simeq C_d\left[1 - |Q_d|\left(\frac{\omega}{C_d}\right)^2(\cos\chi + 3\xi_d\sin\chi)\right], \qquad (42)$$

$$\widetilde{\xi}_d \simeq \xi_d - |Q_d|\left(\frac{\omega}{C_d}\right)^2[\sin\chi - 2\xi_d\cos\chi]$$

with

$$C_d \simeq \sqrt{\frac{\gamma P^e}{\alpha_\infty\rho^e}}, \quad \xi_d \sim \sqrt{\frac{\omega_c}{\omega}}.$$

In the whole frequency range and for common pore morphology, the estimates of the tensors (Sec. III C) suggest the following assessment, where $\mathcal{G}$ is a dimensionless geometric factor, which accounts for the pore geometry and the ratio between pore and elementary representative volume sizes:

$$Q(-ih)^2 \sim \mathcal{G}\left[\frac{\mathcal{K}}{^0K}\right]\frac{i\omega\mu}{\gamma P^e},$$

i.e.,

$$\frac{C}{\widetilde{C}}\frac{1 - i\widetilde{\xi}}{1 - i\xi} - 1 \sim \mathcal{G}\left[\frac{\mathcal{K}}{^0K}\right]\frac{i\omega\mu}{\gamma P^e}.$$

C. Diffraction and critical frequency

Since the diffraction effects are $O(\epsilon^2)$ they are actually significant when $\epsilon \rightarrow 1$, i.e., for frequencies close to ω_d. To identify the qualitative nature of the diffraction when it becomes significant, it is necessary to compare ω_d to ω_c.

According to the properties of P_2 waves, the wavelength may be assessed as (where $C_a = \sqrt{\gamma P^e/\rho^e}$ is the sound velocity in air)

$$\omega < \omega_c \Rightarrow \lambda/2\pi \approx \frac{C_a}{\sqrt{\omega\omega_c}}, \quad \omega > \omega_c \Rightarrow \lambda/2\pi \approx \frac{C_a}{\omega\sqrt{\alpha_\infty}}.$$

Applied at the diffraction frequency (ω_d is such that $\lambda/2\pi = l$), these relations give

$$\frac{\omega_d}{\omega_c} < 1 \Rightarrow \frac{\omega_d}{\omega_c} = \left(\frac{C_a}{l\omega_c}\right)^2,$$

$$\frac{\omega_d}{\omega_c} > 1 \Rightarrow \frac{\omega_d}{\omega_c} = \frac{C_a}{l\omega_c\sqrt{\alpha_\infty}} = \frac{C_a\rho^e}{\mu}\frac{\mathcal{K}}{\phi l}.$$

If we now consider:

(1) Air properties at the normal conditions, $\mu/C_a\rho^e \simeq 1.810^{-5}$ Pa s / 343 m/s1.2 kg/m$^3 \simeq 4, 3.10^{-8}$ m and
(2) The very rough permeability estimate $\mathcal{K} = O(\phi l^2)$, it appears that as soon as $l > 4 \times 10^{-8}$ m, then $\omega_d > \omega_c$. This

Claude Boutin: Rayleigh scattering of poroacoustic waves

means that the diffraction effect is significant in the dynamic regime, except for materials of pores smaller than about 50 nm.

The self-consistent approach for the intrinsic permeability of granular material—of characteristic size $l=2R$ and porosity ϕ (Boutin, 2000)—supplies the more realistic estimate:

$$\mathcal{K} = R^2 \Psi(\phi),$$

$$\Psi(\phi) = \frac{1}{3}\left[\frac{2+3(1-\phi)^{5/3}}{3+2(1-\phi)^{5/3}}\frac{1}{(1-\phi)^{1/3}}-1\right].$$

Using this expression, Fig. 2 shows the ratio ω_d/ω_c versus the porosity for several realistic characteristic sizes. Clearly, the diffraction takes place in the viscous regime only for materials presenting rather fine pores (say less than 1 μm) and/or small porosity. For most of the sound-absorbing materials, the diffraction frequency is to be expected within the dynamic regime.

V. A SIMPLE ACADEMIC EXAMPLE

To illustrate these results we consider a periodic array of parallel slits, Fig. 3. The porous medium consists in parallel impervious rigid plane plates of normal vector e_3, of negligible thickness (i.e., $\phi \simeq 1$), separated by gas layers of constant thickness $2a$. This medium is isotropic in the plane (e_1,e_2) and impervious in direction e_3. The one-dimensional geometry enables the exact determination (detailed in Appendix C) of the local fields and macroscopic tensors. For the sake of simplicity, the adiabatic regime is assumed. From the
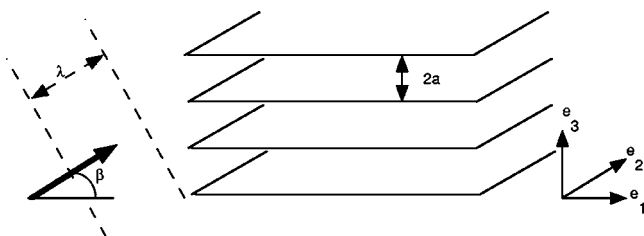
FIG. 3. Porous media made of a periodic array of slits.

"cell," symmetry $^1K=0$ and the first corrector ^1V cancel out. According to Sec. III C 6 the description is, up to the second order, as follows.

Zero order:

$$\mathrm{div}(^0V) + i\omega\phi\frac{^0P}{\gamma P^e} = 0, \quad \mu^0V = -\,^0K.\nabla^0P \tag{43}$$

Second order (*with adiabatic tensor* ^2N):

$$\begin{cases} \mathrm{div}(^2V) + i\omega\phi\dfrac{^2P}{\gamma P^e} = 0, \\ \mu^2V = -\,^0K.\nabla^2P - \,^2K.\nabla\nabla\nabla^0P - \dfrac{i\omega\mu}{P^e}\,^2N.\nabla^0P \end{cases}.$$

A. Macroscopic tensors

Using the notations, (recall that $\delta_v = \sqrt{\mu/i\omega\rho^e}$)

$$u = \frac{a}{\delta_v}, \quad C = \cosh(u), \quad S = \sinh(u), \quad T = \tanh(u),$$

the zero-order dynamic permeability tensor takes the classical diagonal form ($^0K_i^j = 0$ for $i \neq j$):

$$^0K_1^1 = {}^0K_2^2 = \delta_v^2\left[1 - \frac{T}{u}\right], \quad {}^0K_3^3 = 0.$$

At the next order, as expected from the cell symmetry calculations show that ^1K vanishes, whereas some local fields ^1k are not zero:

$$\forall i,j,k, \quad {}^1K_i^{jk} = 0.$$

After the algebra described in Appendix C, the nonvanishing components of ^2K are:

$$^2K_1^{111} = {}^2K_2^{222} = \delta_v^4\left[\frac{3}{2} - \frac{5T}{2u} - \left(\frac{1}{6} - \frac{1}{u^2}\right)T^2\right],$$

$$^2K_1^{122} = {}^2K_2^{211} = \delta_v^4\left[\frac{3}{2} - \frac{3T}{2u} - \frac{T^2}{2}\right],$$

$$^2K_2^{112} = {}^2K_1^{221} = \delta_v^4\left[-\frac{T}{u} + \left(\frac{1}{3} + \frac{1}{u^2}\right)T^2\right],$$

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Claude Boutin: Rayleigh scattering of poroacoustic waves 1897

$$^2K_1^{133} = {}^2K_2^{233} = \delta_v^4\left[\frac{1}{2} - \frac{T}{2u} - \frac{T^2}{2} - \frac{u}{SC}\right],$$

$$^2K_3^{311} = {}^2K_3^{322} = \delta_v^4\left[-1 - \frac{u^2}{3} + \frac{u}{T}\right],$$

$$^2K_3^{131} = {}^2K_3^{232} = {}^2K_1^{313} = {}^2K_2^{323} = \delta_v^4\left[-1 + \frac{2T}{u} - \frac{u}{SC}\right],$$

$$^2K_3^{113} = {}^2K_3^{223} = {}^2K_1^{331} = {}^2K_2^{332} = \delta_v^4\left[-1 + \left(\frac{1}{u} + \frac{u}{3}\right)T\right].$$

Finally, tensor ^2N is diagonal ($^2N_i^j = 0$ for $i \neq j$) and has the following components:

$$^2N_1^1 = {}^2N_2^2 = \frac{{}^0K_1^1}{\gamma}, \quad {}^2N_3^3 = \frac{a^2}{3} = \frac{\mathcal{K}}{\gamma}.$$

B. Plane wave diffraction

Consider now a plane wave propagating in a direction d inclined at an angle β with the slits, i.e., $d = \cos(\beta)e_1 + \sin(\beta)e_3$. In this direction, the wave number h_d is given by

$$(-ih_d)^2 = \frac{i\omega}{\gamma P^e}\frac{\mu}{{}^0K_1^1\cos(\beta)^2}$$

and the (adiabatic) coefficient Q_d takes the form, see Eq. (38):

$$2Q_d = \frac{{}^2K_1^{111}\cos(\beta)^2 + ({}^2K_1^{133} + {}^2K_3^{311} + 2{}^2K_1^{313} + 2{}^2K_1^{331})\sin(\beta)^2}{{}^0K_1^1} + {}^0K_1^1\cos(\beta)^2 + \mathcal{K}\sin(\beta)^2.$$

Thus the dimensionless corrective term $Q_d(ih_d)^2$ reads:

$$Q_d(-ih_d)^2 = \frac{1}{2}\left\{1 + \frac{{}^2K_1^{111}}{({}^0K_1^1)^2} + \tan(\beta)^2\left[\frac{\mathcal{K}}{{}^0K_1^1}\right.\right.$$
$$\left.\left. + \frac{{}^2K_1^{133} + {}^2K_3^{311} + 2{}^2K_1^{313} + 2{}^2K_1^{331}}{({}^0K_1^1)^2}\right]\right\}\frac{i\omega\mu}{\gamma P^e}.$$

1. Propagation parallel to the slits

When focusing on waves propagating parallel to the slits, the corrector reduces to

$$Q_{e_1}(ih_{e_1})^2 = \frac{1}{2}\left[1 + \frac{{}^2K_1^{111}}{({}^0K_1^1)^2}\right]\frac{i\omega\mu}{\gamma P^e}.$$

From expressions of $^0K_1^1$ and $^2K_1^{111}$, it can be shown that, all over the whole frequency range, $^2K_1^{111}({}^0K_1^1)^{-2}$ is very close to a constant real value (varying from $7/5$ at low frequency to $4/3$ at high frequency). Consequently:

(1) At low frequency, Q_{e_1} tends to the real value $2a^2/5 = 6\mathcal{K}/5$ and, according to Eq. (41), the scattering perturbation is near to that of diffusion waves.
(2) At high frequency, $Q_{e_1} \sim -i/\omega$: This leads to *atypical* effects, i.e., [cf. Eq. (42)], a velocity dispersion varying according to $\sqrt{\omega}$, and an additional attenuation varying according to ω.

2. Oblique propagation

In oblique directions a significant anisotropic effect is induced in the whole frequency range by the magnification coefficient $\tan(\beta)^2$. Further,

(1) At low frequency, Q_d tends to a real constant ($\sim\mathcal{K}$) and therefore diffraction will be similar to that of diffusion waves.
(2) At high frequency, terms that vary according to $i\omega$ exactly compensate each other. Thus, $Q_d \sim \sqrt{i\omega}$, implying that $Q_d(ih_d)^2$ varies as $\omega^2\sqrt{i\omega}$. Again, the consequences of diffraction differ from that of elastic waves: Both velocity dispersion and additional attenuation vary according to $\omega^2\sqrt{i\omega}$.

This simple case shows that the diffraction correction presents a rather complex frequency dependence. Figures 4–6 depict the strong quantitative and qualitative differences when (i) The diffraction frequency lies within the viscous, critical, or inertial regime (those cases are reached by varying the thickness of the gas layer) and (ii) the propagation is parallel to the slit or in oblique direction. It is worth noting that the atypical results at high frequency are strongly related to the one-dimensional-geometry and should not be generalized to other porous materials.

VI. CONCLUSION

To investigate the long wave scattering in air-saturated porous media, the periodic homogenization method was applied, extending the analysis up to the second-order terms. The actual enriched second-order description given by Eqs. (20), (30), and (31) accounts for the viscous, thermal, and inertia effects at the pore scale and is valid even for large concentration of solid.

In the Rayleigh domain, the improving of the continuum description implies a nonlocal behavior (associated with microstructural sources) and specific tensors depending on the microstructure and the frequency. It is shown that—up to the second order at least—the tensors associated the viscoinertial

1898   J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Claude Boutin: Rayleigh scattering of poroacoustic waves

FIG. 4. Diffraction effect within the viscous regime. The thickness $2a=5\times10^{-8}$ m gives a diffraction frequency $\omega_d\simeq0.01\omega_c$, $(f_c\simeq100$ GHz). The relative variation of velocity, $\Delta C/C=(\widetilde{C}-C)/C$, (plain line) and variation of attenuation, $\Delta\xi=\widetilde{\xi}-\xi$, (dashed line) are presented vs $\mathrm{Log}[\omega/\omega_c]$, for propagation parallel to the slits (tick line) and in oblique direction $\beta=\pi/4$ (normal line). Velocity and attenuation significantly increase, similarly in both directions.

FIG. 5. Diffraction effect within the viscoinertial regime. The thickness $2a=5\times10^{-6}$ m gives a diffraction frequency $\omega_d\simeq\omega_c$, $(f_c\simeq10$ MHz). Same legend as in Fig. 4. Wave parallel to the slits presents an increase of attenuation and a slight increase of velocity. Oblique wave becomes more attenuated with a transition from a slight increase of velocity to a decrease.
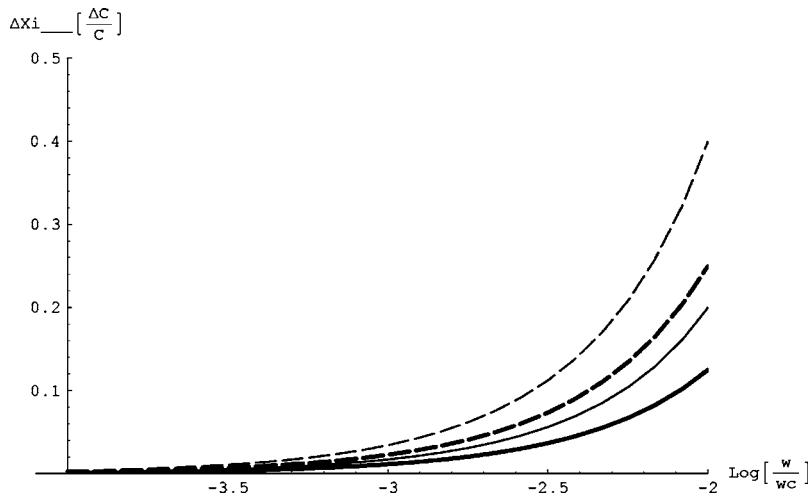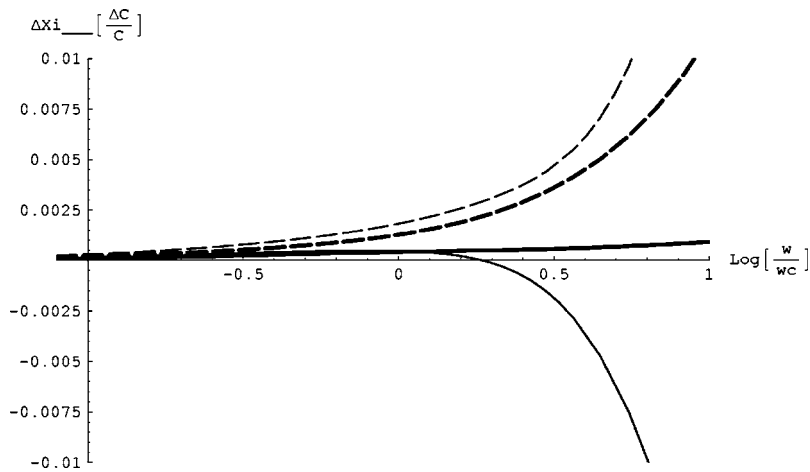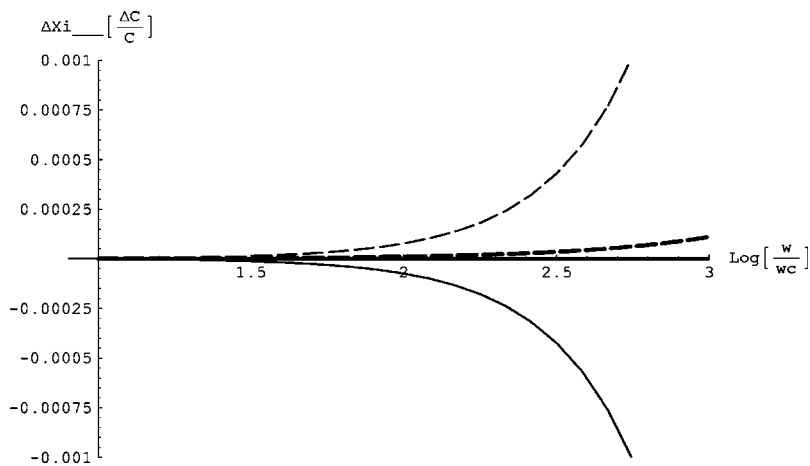

FIG. 6. Diffraction effect within the inertial regime. The thickness $2a=5\times10^{-4}$ m gives a diffraction frequency $\omega_d\simeq10^3\omega_c$, $(f_c\simeq1$ kHz). Same legend as in Fig. 4. Diffraction makes oblique waves more attenuated and of slower velocity, while parallel waves are almost unaffected.

effects can be determined separately, while the thermal effects become coupled with viscous effects. A key result is that the first-order macropressure is demonstrated to vanish and that the correctors are reported at the second order. This points out the robustness of the usual continuum description, and explains in some manner the abrupt transition from the behavior in the Rayleigh domain to the behavior at higher frequencies.

The perturbation of plane wave propagation is derived from this description and interpreted as interferences between the zero-order field and the field—coherent and linearly amplified by a *self-resonance* mechanism—built up from the diffracted sources. The frequency dependence of the velocity's and attenuation's correction is rather complex due to the dispersion of the P_2 wave. Effects close to that observed for diffusion waves are recovered at low frequency. At high frequency an accurate discussion of these theoretical results would require numerical computation of high order tensors. An example is presented for the simple case of an array of slits. In that case, the atypical effect in high frequency range is strongly anisotropic and differs from what is observed with purely elastic waves.

Let us underline that those results are only valid in the scale separation frequency range. At higher frequencies, this approach becomes irrelevant since the scattered waves radiate in other directions than the initial wave. Another limitation is the periodicity assumption. Nevertheless, it may be inferred that this hypothesis does not modify the nature of the results for long wave in random media, provided that they can be characterized by a representative volume smaller than the wavelength. However, regarding shorter waves, strong divergences have to be expected, mostly because the diffracted field in random media loses its coherence.

APPENDIX A: HOMOGENIZATION PROCESS UP TO THE SECOND ORDER

This appendix gives the complement of the homogenization process not detailed in the main text. To save notations, the Einstein convention is used, the derivative $\partial f / \partial y_i$ is denoted by $f_{,i}$, the high indexes of the particular solutions are related to the forcing term, the low indexes indicate the components of the field and derivative, e.g., $v = v_i . e_i$; $\mathrm{div}_y(v) = v_{i,i}$; δ_i^m is the Kronecker symbol ($\delta_i^m = 1$ if $i = m$; $= 0$

otherwise). The average on the pores domain is denoted by $\langle - \rangle = (1/\Omega_f) \int_{\Omega_f} - \, d\Omega$.

1. The leading and first order

These two first steps are detailed in Sec. III B. The differential systems $\{S_t^0\}$ satisfied by the particular solutions $^0\pi$, is

$$S_t^0(^0\pi) \begin{cases} i\omega \rho^e c_p \, ^0\pi - \kappa \Delta_y(^0\pi) = i\omega \dfrac{P^e}{T^e}, \\[2mm] ^0\pi_{/\Gamma} = 0, \, ^0\pi \ \ \Omega\text{-periodic.} \end{cases}$$

Keeping in mind that $P^e / T^e = \rho^e c_p (1 - 1/\gamma)$, it clearly appears that $^0\pi$ depends on the frequency throughout the single parameter $\delta_t^{-2} = i\omega \rho^e c_p / \kappa$. Thus, $<^0\pi> = {}^0\Pi$, is a function of the thermal dimensionless frequency ω/ω_t [ω_t being defined by Eq. (22)].

As for the solutions $(-^0k^m/\mu, ^1\alpha^m)$, they are governed by the set $\{S_v^0\}$:

$$S_v^0(^0k^m, ^1\alpha^m) \begin{cases} i\omega \dfrac{\rho^e}{\mu} \, ^0k_i^m - \Delta_y(^0k_i^m) - {}^1\alpha_{,i}^m = \delta_i^m, \\[2mm] ^0k_{i,i}^m = 0, \\[2mm] ^0k_{/\Gamma}^m = 0, \quad \langle ^1\alpha^m \rangle = 0, \\[2mm] ^0k^m, ^1\alpha^m \ \ \Omega\text{-periodic.} \end{cases}$$

This form points out that $(-^0k^m, ^1\alpha^m)$ depends on the frequency throughout the single parameter $\delta_v^{-2} = i\omega \rho^e / \mu$. Therefore, $\phi\langle ^0k \rangle = {}^0K$ is function of the viscous dimensionless frequency ω/ω_c [ω_c being defined by Eq. (21)].

The differential system $\{S_t^1\}$ satisfied by the particular solutions $^1\pi^m$ is given in the following:

$$S_t^1(^1\pi^m) \begin{cases} i\omega \rho^e c_p \, ^1\pi^m - \kappa \Delta_y(^1\pi^m) = i\omega \dfrac{P^e}{T^e} \, ^1\alpha^m + 2\kappa \, ^0\pi_{,m}, \\[2mm] ^1\pi_{/\Gamma} = 0, \quad ^1\pi \ \ \Omega\text{-periodic.} \end{cases}$$

The presence of $^1\alpha^m$ implies a viscothermal coupling. Then $\langle ^1\pi \rangle = {}^1\Pi$ depends on both dimensionless frequencies ω/ω_t and ω/ω_c.

2. The second order

The gas flow problem $\{S_v^1\}$

The set of equations $\{S_v^1\}$ governing $(^1v, {}^2p)$ reads explicitly:

$$S_v^1 \begin{cases} -i\omega \rho^e \, ^1v + \Delta_y(\mu \, ^1v) - \nabla_y \, ^2p = \nabla_x \, ^1p - 2\Delta_{yx}(\mu \, ^0v) + i\omega \mu \nabla_y \dfrac{^0\theta}{T^e}, \\[2mm] \mathrm{div}_y(^1v) = -\mathrm{div}_x(^0v) + \mathrm{div}_x(\langle ^0v \rangle) + i\omega \left[\dfrac{^0\theta - \langle ^0\theta \rangle}{T^e} \right], \\[2mm] v_{/\Gamma}^1 = 0, \quad ^1v, \text{ and } ^1p \, \Omega\text{-periodic.} \end{cases}$$

Claude Boutin: Rayleigh scattering of poroacoustic waves

Note the following:

(1) Taking advantage of $\langle \mathrm{div}_y(^1v)\rangle=0$ (see Sec. III A) the average mass balance equation has been subtracted from the local mass balance.
(2) The set $\{S_v^1\}$ is identical to the set $\{S_v^0\}$, except for the right-hand side of the equations acting as forcing terms induced by the solutions at previous order.

When substituting Eqs. (17) and (18) of $^0\theta$, ^0v, and ^1p it appears that the forcing terms may be split into:

(1) A term associated with $\nabla_x\,^1\widehat{p}$, which leads to an identical problem than at the preceding order.
(2) Terms coming from ^0v, $\langle^0v\rangle$ and ^1p, associated with the second gradient of the zero-order pressure $\nabla_x\nabla_x{}^0P$.
(3) Terms coming from $^0\theta$ and $\langle^0\theta\rangle$, associated to the zero-order pressure ^0P.

As a consequence of the linearity, the solution is a linear combination of particular solutions associated with each of these forcing terms. Consequently the solution is in the form:

$$^2p(x,y) = {}^2\widehat{p}(x) + {}^1\alpha\,.\,\nabla_x\,{}^1\widehat{p}(x) + {}^2\alpha\,.\,.\,\nabla_x\nabla_x{}^0P(x)$$

$$+\,\frac{i\omega\mu}{P^e}\,{}^2v\,{}^0P(x),\tag{A1}$$

$$-\mu\,{}^1v(x,y) = {}^0k\,.\,\nabla_x\,{}^1\widehat{p}(x) + {}^1k\,.\,.\,\nabla_x\nabla_x{}^0P(x)$$

$$+\,\frac{i\omega\mu}{P^e}\,{}^1n\,{}^0P(x).\tag{A2}$$

The solutions $(-{}^1k^{mn}/\mu,\,{}^2\alpha^{mn})$ corresponding to unit second gradients of pressure, $\nabla_x\nabla_x{}^0P(x)=e_m\otimes e_n$, are governed by

$$S_v^1({}^1k^{mn},\,{}^2\alpha^{mn})$$

$$\times\begin{cases} i\omega\dfrac{\rho^e}{\mu}\,{}^1k_i^{mn} - \Delta_y({}^1k_i^{mn}) - {}^2\alpha_{,i}^{mn} = {}^1\alpha^m\delta_i^n + 2\,{}^0k_{i,n}^m,\\[4pt] {}^1k_{i,i}^{mn} = -\,{}^0k_n^m + \langle{}^0k_n^m\rangle,\\[4pt] {}^1k_{/\Gamma}^{mn} = 0;\langle{}^2\alpha^{mn}\rangle = 0,\\[4pt] {}^1k^{mn} \text{ and } {}^2\alpha^{mn}\ \Omega\text{-periodic.} \end{cases}$$

As previously, the single parameter is δ_v^{-2} so that $\phi\langle{}^1k\rangle={}^1K$, only depends on ω/ω_c.

The solution $(-{}^1n/\mu,\,{}^2v)$ corresponding to a pressure, $^0P(x)=P^e/i\omega\mu$, is driven by

$$S_v^1({}^1n,\,{}^2v)\begin{cases} i\omega\dfrac{\rho^e}{\mu}\,{}^1n_i - \Delta_y({}^1n_i) - {}^2v_{,i} = {}^0\pi_{,i},\\[4pt] {}^1n_{i,i} = {}^0\Pi - {}^0\pi,\\[4pt] {}^1n_{/\Gamma} = 0,\langle{}^2v\rangle = 0,\\[4pt] {}^1n,\ \text{and}\ {}^2v\ \Omega\text{-periodic.} \end{cases}$$

The source terms brought a thermal coupling. Therefore $\langle{}^1n\rangle={}^1N$ depends on ω/ω_t and ω/ω_c.

The heat transfer problem $\{S_t^2\}$

As for the temperature, $^2\theta$ is solution of $\{S_t^2\}$:

$$S_t^2\begin{cases} i\omega\rho^e c_p\,{}^2\theta - \kappa\Delta_y({}^2\theta) = {}^2pi\omega + 2\kappa\Delta_{yx}({}^1\theta) + \kappa\Delta_x({}^0\theta),\\[4pt] \theta_{/\Gamma}^2 = 0,\quad \theta^2\ \Omega\text{-periodic.} \end{cases}$$

Inserting Eqs. (17)–(19) and (A1) giving $^0\theta$, $^1\theta$, and ^2p, one notes that the forcing terms are constituted by the following:

(1) Terms associated with $^2\widehat{p}$ and $\nabla_x\,^1\widehat{p}$, which lead to identical problems than at the two preceding orders.
(2) Terms associated with the second gradient of the zero-order pressure $\nabla_x\nabla_x{}^0P$.
(3) Terms associated with the zero order pressure ^0P.

Again, from the linearity, the solution is a linear combination of particular solutions associated with each forcing term:

$$(P^e/T^e)^2\,\theta(x,y) = {}^0\pi\,{}^2\widehat{p}(x) + {}^1\pi\,.\,\nabla_x\,{}^1\widehat{p}(x)$$

$$+\,{}^2\pi\,.\,.\,\nabla_x\nabla_x{}^0P(x) + \frac{i\omega\mu}{P^e}\,{}^2\zeta\,{}^0P(x).$$

$$\tag{A3}$$

The temperature distributions $^2\pi^{mn}$ solutions of $\{S_t^2\}$ under second gradient of pressure $\nabla_x\nabla_x{}^0P(x)=(P^e/T^e)e_m\otimes e_n$ are determined by

$$S_t^2({}^2\pi^{mn})$$

$$\times\begin{cases} i\omega\rho^e c_p\,{}^2\pi^{mn} - \kappa\Delta_y({}^2\pi^{mn}) = i\omega\dfrac{P^e}{T^e}\,{}^2\alpha^{mn} + 2\kappa\,{}^1\pi_{,n}^m + \kappa\,{}^0\pi\delta_n^m,\\[4pt] {}^2\pi_{/\Gamma}^{mn} = 0,\quad {}^2\pi^{mn}\ \Omega\text{-periodic.} \end{cases}$$

The temperature distribution $^2\zeta$ solution of $\{S_t^2\}$ under pressure $^0P(x)=P^e/i\omega\mu$ is determined by

$$S_t^2({}^2\zeta)\begin{cases} i\omega\rho^e c_p\,{}^2\zeta - \kappa\Delta_y({}^2\zeta) = i\omega\dfrac{P^e}{T^e}\,{}^2v,\\[4pt] {}^2\zeta_{/\Gamma} = 0,\quad {}^2\zeta\ \Omega\text{-periodic.} \end{cases}$$

Note that the source terms introduce a viscous coupling in the first case, and a thermoviscous coupling in the second case. Then $^2\Pi$ and ^2Z depend on ω/ω_t and ω/ω_c.

3. The third order: The gas flow problem $\{S_v^2\}$

Finally, the set $\{S_v^2\}$ governing ^3p and ^2v is

$$S_v^2 \begin{cases} -i\omega\rho^{e2}v + \Delta_y(\mu^2 v) - \nabla_y{}^3 p = \dots \\[6pt] = \nabla_x{}^2 p - 2\Delta_{yx}(\mu^1 v) - i\omega\mu\left(\nabla_y\dfrac{^1p}{P^e} - \nabla_y\dfrac{^1\theta}{T^e}\right) - \Delta_x(\mu^0 v) - i\omega\mu\left(\nabla_x\dfrac{^0P}{P^e} - \nabla_x\dfrac{^0\theta}{T^e}\right) \\[10pt] \text{div}_y(^2v) = -\text{div}_x(^1v) + \text{div}_x(\langle ^1v\rangle) - i\omega\left[\dfrac{^1p - \langle ^1p\rangle}{P^e} - \dfrac{^1\theta - \langle ^1\theta\rangle}{T^e}\right] \\[10pt] v_{/\Gamma}^2 = 0, \quad ^2v \text{ and } ^3p \ \Omega\text{-periodic.} \end{cases}$$

Following the same reasoning as mentioned earlier, one shows that the solution reads:

$$^3p(x,y) = ^3\widehat{p}(x) + {}^1\alpha \cdot \nabla_x{}^2\widehat{p}(x) + {}^2\alpha \cdot \cdot \nabla_x\nabla_x{}^1\widehat{p}(x)$$
$$+ {}^3\alpha \dots \nabla_x\nabla_x{}^0P(x)$$
$$+ \frac{i\omega\mu}{P^e}[^2\nu{}^1\widehat{p}(x) + {}^3\nu\nabla_x{}^0P(x)],$$

$$-\mu^2 v(x,y) = ^0k \cdot \nabla_x{}^2\widehat{p}(x) + {}^1k \cdot \cdot \nabla_x\nabla_x{}^1\widehat{p}(x)$$
$$+ {}^2k \cdot \cdot \nabla_x\nabla_x{}^0P(x)$$
$$+ \frac{i\omega\mu}{P^e}[^1n{}^1\widehat{p}(x) + {}^2n \cdot \nabla_x{}^0P(x)].$$

The solutions $(-^2k^{mnp}/\mu, {}^3\alpha^{mnp})$ corresponding to unit third gradient of pressure, $\nabla_x\nabla_x\nabla_x{}^0P(x) = e_m \otimes e_n \otimes e_p$ are governed by

$$S_v^2(^2k^{mnp}, {}^3\alpha^{mnp})$$
$$\times \begin{cases} i\omega\dfrac{\rho^e}{\mu}{}^2k_i^{mnp} - \Delta_y(^2k_i^{mnp}) - {}^3\alpha_{,i}^{mnp} \\[6pt] \quad = {}^2\alpha^{mn}\delta_i^p + 2{}^1k_{i,p}^{mn} + {}^0k_i^m\delta^{np}, \\[6pt] ^2k_{i,i}^{mnp} = -{}^1k_p^{mn} + \langle ^1k_p^{mn}\rangle, \\[6pt] ^2k_{/\Gamma}^{mnp} = 0, \quad \langle ^3\alpha^{mnp}\rangle = 0, \\[6pt] ^2k^{mnp}, \quad ^3\alpha^{mnp} \ \Omega\text{-periodic.} \end{cases}$$

Once more, the single physical parameter is δ_v^{-2} so that $\phi\langle ^2k\rangle = {}^2K$, depends on ω/ω_c only.

The solutions $(-^2n^l/\mu, {}^3\nu^l)$ corresponding to gradients of pressure $\nabla_x{}^0P(x) = (P^e/i\omega\mu)e_l$ are governed by

$$S_v^2(^2n^l, {}^3\nu^l)$$
$$\times \begin{cases} i\omega\dfrac{\rho^e}{\mu}{}^2n_i^l - \Delta_y(^2n_i^l) - {}^3\nu_{,i}^l = {}^1\alpha_{,i}^l - {}^1\pi_{,i}^l + \delta_i^l - {}^0\pi\delta_i^l, \\[6pt] ^2n_{i,i}^l = {}^1\alpha^l - {}^1\pi^l - \langle ^1\pi^l\rangle, \\[6pt] ^2n_{/\Gamma}^l = 0; \quad \langle ^3\nu^l\rangle = 0, \\[6pt] ^2n^l \text{ and } ^3\nu^l \ \Omega\text{-periodic.} \end{cases}$$

The source terms brought viscous and thermoviscous coupling, and then $\langle ^2n\rangle = {}^2N$ depends on ω/ω_t and ω/ω_c.

The existence and uniqueness of the solutions of all these problems can be demonstrated from their variational formulation, not developed here.

APPENDIX B: TENSORS PROPERTIES

The leading idea to establish the properties of the macroscopic tensors is based on energy reciprocity. The principle is to calculate separately the virtual energies of a solution under another and reciprocally. Then, comparing the two expressions, and using the fact that the virtual kinetic and dissipated energy (or inertial and conduction energy for thermal problems) are identical in both cases, tensor identities are derived.

As a first example, take the scalar product of the momentum balance of the solution $^0k^m$ [set $S_v^0(^0k^m, {}^1\alpha^m)$, Appendix A] with the solution $^0k^n$ and integrate on the pore volume. It gives

$$\langle -{}^1\alpha_{,i}^{m0}k_i^n\rangle + i\omega\frac{\rho^e}{\mu}\langle ^0k^m \cdot {}^0k^n\rangle - \langle \Delta_y(^0k_i^m){}^0k_i^n\rangle = \langle ^0k_m^n\rangle.$$

From the divergence theorem, the first term may be transformed as

$$\langle -{}^1\alpha_{,i}^{m0}k_i^n\rangle = \langle ^1\alpha^{m0}k_{i,i}^n\rangle - \frac{1}{\Omega_f}\int_{\partial\Omega_f}{}^1\alpha^{m0}k^n \cdot n\,ds = 0$$

[because of, (i) the zero divergence of $^0k^n$ and (ii) the adherence and the periodicity conditions]. Similarly, from the divergence theorem

$$\langle \Delta_y(^0k_i^m){}^0k_i^n\rangle = \langle ^0k_{i,jj}^m{}^0k_i^n\rangle = \langle (^0k_{i,j}^m{}^0k_i^n)_{,j} - {}^0k_{i,j}^m{}^0k_{i,j}^n\rangle$$
$$= \frac{1}{\Omega_f}\int_{\partial\Omega_f}(\nabla_y{}^0k^m \cdot n) \cdot {}^0k^n ds$$
$$- \langle \nabla_y{}^0k^m \cdot \cdot \nabla_y{}^0k^n\rangle$$

and the surface integral vanishes because of the adherence and periodicity. Thus finally:

$$i\omega\frac{\rho^e}{\mu}\langle ^0k^m \cdot {}^0k^n\rangle + \langle \nabla_y(^0k^m) \cdot \cdot \nabla_y(^0k^n)\rangle = \langle ^0k_m^n\rangle.$$

Inverting the role of $^0k^m$ and $^0k^n$ and comparing the two expressions lead directly to

$$\langle ^0k_m^n\rangle = \langle ^0k_n^m\rangle,$$

which demonstrate the symmetry of the dynamic permeability ^0K, [Auriault, 1980].

1. Skew symmetry of ^1K

Apply now this procedure to $^0k^m$ and $^1k^{np}$. Consider first the virtual work of $^0k^m$ under $^1k^{np}$:

$$\langle -\,^1\alpha_{,i}^{m1}k_i^{np}\rangle + i\omega\frac{\rho^e}{\mu}\langle ^0k^m\,.\,^1k^{np}\rangle - \langle\Delta_y(^0k_i^m)^1k_i^{np}\rangle = \langle ^1k_m^{np}\rangle.$$

Using the same argument as above, and taking into account (i) the expression of the divergence of $^1k^{np}$ [set $S_v^1(^1k^{mn},\,^1\alpha^{mn})$, Appendix A] and (ii) $\langle ^1\alpha^m\rangle=0$, the first term becomes

$$\langle -\,^1\alpha_{,i}^{m1}k_i^{np}\rangle = \langle ^1\alpha^{m1}k_{i,i}^{np}\rangle + 0 = \langle ^1\alpha^m(-\,^0k_p^n + \langle ^0k_p^n\rangle)\rangle =$$
$$- \langle ^1\alpha^{m0}k_p^n\rangle$$

For the same reasons as previously:

$$- \langle\Delta_y(^0k_i^m)^1k_i^{np}\rangle = \langle\nabla_y(^0k^m)\,.\,.\,\nabla_y(^1k^{np})\rangle$$

and thus

$$- \langle ^1\alpha^{m0}k_p^n\rangle + i\omega\frac{\rho^e}{\mu}\langle ^0k^m\,.\,^1k^{np}\rangle + \langle\nabla_y(^0k^m)\,.\,.\,\nabla_y(^1k^{np})\rangle$$
$$= \langle ^1k_m^{np}\rangle. \tag{B1}$$

In a second step, consider the virtual work of $^1k^{np}$ under $^0k^m$. From the momentum balance [cf. set $S_v^1(^1k^{mn},\,^1\alpha^{mn})$, Appendix A] we get

$$\langle -\,^2\alpha_{,i}^{np0}k_i^m\rangle + i\omega\frac{\rho^e}{\mu}\langle ^1k^{np}\,.\,^0k^m\rangle - \langle\Delta_y(^1k_i^{np})^0k_i^m\rangle$$
$$= \langle ^1\alpha^{n0}k_p^m\rangle + 2\langle ^0k_{i,p}^n\,^0k_i^m\rangle. \tag{B2}$$

Here again, the first term is null because of the zero divergence of $^0k^m$, and $-\langle\Delta_y(^1k_i^{np})^0k_i^m\rangle$ may be changed into $\langle\nabla_y(^0k^m)\,.\,.\,\nabla_y(^1k^{np})\rangle$. Then, subtracting Eqs. (B1) and (B2) one obtains

$$- \langle ^1\alpha^{m0}k_p^n\rangle + \langle ^1\alpha^{n0}k_p^m\rangle + 2\langle ^0k_{i,p}^n\,^0k_i^m\rangle = \langle ^1k_m^{np}\rangle. \tag{B3}$$

Note that this expression enables one to calculate $\langle ^1k_m^{np}\rangle$ from the solutions at previous order *without calculating* $^1k_m^{np}$. Moreover, inverting m and n gives

$$\langle ^1k_m^{np}\rangle + \langle ^1k_n^{mp}\rangle = 2(\langle ^0k_{i,p}^n\,^0k_i^m\rangle + \langle ^0k_{i,p}^m\,^0k_i^n\rangle) = \cdots$$
$$= 2\langle(^0k^n\,.\,^0k^m)_{,p}\rangle = -\frac{2}{\Omega_f}\int_{\partial\Omega_f}[^0k^n\,.\,^0k^m]e_p\,.\,nds = 0$$

(because of periodicity and adherence conditions). This demonstrates the skew symmetry of ^1K relative to the two first indexes:

$$^1K_m^{np} = -\,^1K_n^{mp}.$$

2. Relation between ^1N and $^1\Pi$

Similarly an expression of ^1N is derived by considering the following.

(1) The virtual work of $^0k^m$ under ^1n:

$$\langle -\,^1\alpha_{,i}^{m1}n_i\rangle + i\omega\frac{\rho^e}{\mu}\langle ^0k^m\,.\,^1n\rangle - \langle\Delta_y(^0k_i^m)^1n_i\rangle = \langle ^1n_m\rangle,$$

which reads after transformation using the mass balance of $S_v^1(^1n,\,^2\nu)$, Appendix A:

$$\langle ^1\alpha^m(^0\Pi - \,^0\pi)\rangle + i\omega\frac{\rho^e}{\mu}\langle ^0k^m\,.\,^1n\rangle + \langle\nabla_y\,^0k^m\,.\,.\,\nabla_y\,^1n\rangle$$
$$= \langle ^1n_m\rangle.$$

(2) The virtual work of ^1n under $^0k^m$ [cf. the momentum balance of $S_v^1(^1n,\,^2\nu)$, Appendix A]:

$$i\omega\frac{\rho^e}{\mu}\langle ^1n\,.\,^0k^m\rangle - \langle\Delta_y(^1n_i)^0k_i^m\rangle - \langle ^2\nu_{,i}\,^0k_i^m\rangle = \langle ^0\pi_{,i}\,^0k_i^m\rangle,$$

which reduces to

$$i\omega\frac{\rho^e}{\mu}\langle ^0k^m\,.\,^1n\rangle + \langle\nabla_y\,^0k^m\nabla_y\,^1n\rangle = 0.$$

Reporting this result in the first equality gives

$$\langle ^1n_m\rangle = -\langle ^1\alpha^{m0}\pi\rangle.$$

Independently, an expression of $^1\Pi$ is derived by considering:

(1) The virtual work of $^0\pi$ under $^1\pi^m$, cf. set $S_t^0(^0\pi)$, Appendix A:

$$\langle i\omega\rho^e c_p\,^0\pi^1\pi^m\rangle - \langle\kappa\Delta_y(^0\pi)^1\pi^m\rangle = i\omega\frac{P^e}{T^e}\langle ^1\pi^m\rangle.$$

(2) The virtual work of $^1\pi^m$ under $^0\pi$, cf. set $S_t^1(^1\pi^m)$, Appendix A:

$$\langle i\omega\rho^e c_p\,^1\pi^{m0}\pi\rangle - \langle\kappa\Delta_y(^1\pi^m)^0\pi\rangle = i\omega\frac{P^e}{T^e}\langle ^1\alpha^{m0}\pi\rangle$$
$$+ 2\langle\kappa\,^0\pi_{,m}\,^0\pi\rangle.$$

Noting that

$$2\kappa\langle ^0\pi_{,m}\,^0\pi\rangle = \kappa\langle(^0\pi)_{,m}^2\rangle = -\frac{2\kappa}{\Omega_f}\int_{\partial\Omega_f}(^0\pi)^2e_p\,.\,nds = 0$$

(because of the periodicity and isotherm boundary conditions), one deduces after usual transformations:

$$\langle ^1\alpha^{m0}\pi\rangle = \langle ^1\pi^m\rangle.$$

Comparing with the above expression of $\langle ^1n_m\rangle$, one obtains

$$\langle ^1n_m\rangle = -\langle ^1\pi^m\rangle, \text{ i.e., } ^1N_m = -\phi^1\Pi^m.$$

3. Expression and symmetry of ^2K

Pursuing the analysis for ^2K, two identities can be established from:

(1) The virtual work of $^0k^m$ under $^2k^{npq}$ and of $^2k^{npq}$ under $^0k^m$:

$$\langle ^2k_m^{npq}\rangle = \langle ^0k^n\,.\,^0k^m\rangle\delta^{pq} - \langle ^1\alpha^{m1}k_q^{np}\rangle + \langle ^2\alpha^{np0}k_q^m\rangle$$
$$- 2\langle ^1k_i^{np0}k_{i,q}^m\rangle.$$

(2) The virtual work of $^1k^{np}$ under $^1k^{mq}$:

$$-\langle^2\alpha^{mq0}k^{np}\rangle + \langle\nabla_y(^1k^{mq})\,.\,.\,\nabla_y(^1k^{np})\rangle + i\omega\frac{\rho^e}{\mu}\langle^1k^{mq}\,.\,^1k^{np}\rangle$$

$$=\langle^1\alpha^{m1}k^{np}\rangle + 2\langle^1k^{np0}_i k^m_{i,q}\rangle.$$

Combining these relations leads to

$$\langle^2k^{npq}_m\rangle = \langle^0k^n\,.\,^0k^m\rangle\delta^{pq} + \langle^2\alpha^{mq0}k^n_p\rangle + \langle^2\alpha^{np0}k^m_q\rangle\cdots$$

$$-\left(\langle\nabla_y(^1k^{mq})\,.\,.\,\nabla_y(^1k^{np})\rangle + i\omega\frac{\rho^e}{\mu}\langle^1k^{mq}\,.\,^1k^{np}\rangle\right),$$

which provides the demonstration of the symmetry of ^2K relative to the two pairs of indexes:

$$^2K^{npq}_m = {}^2K^{mqp}_n.$$

4. Expression of ^2N and ^2Z

As for ^2N, the comparison of the virtual works of $^0k^m$ under $^2n^n$ and reciprocally gives

$$^2N^n_m/\phi = \langle^2n^n_m\rangle = \langle^0k^m_n(1 - {}^0\pi)\rangle + \langle^1\alpha^m(^1\alpha^n - {}^1\pi^n)\rangle.$$

Finally, the virtual works of $^0\pi$ under $^2\zeta$ and inversely lead to

$$^2Z = \langle^2\zeta\rangle = \langle^0\pi^2\nu\rangle.$$

APPENDIX C: HIGH ORDER TENSOR FOR A PERIODIC ARRAY OF SLITS

Owing to the one-dimensional geometry, the fields depend locally on variable y_3 only, and the axes e_1 and e_2 play the same role. In this context, the problem $\{S^0_v\}$ governing the dynamic permeability is rewritten in the simple form:

$$\begin{cases} i\omega\frac{\rho^e}{\mu}{}^0k^m_i - {}^0k^m_{i,33} - {}^1\alpha^m_{,3} = \delta^m_i, \\ {}^0k^m_{3,3} = 0, \\ {}^0k^m(\pm a) = 0, \quad \langle^1\alpha^m\rangle = 0. \end{cases}$$

The resolution is straightforward and gives, where $\delta_v = \sqrt{(\mu/i\omega\rho^e)}$; $y^* = y_3/\delta_v$; $u = a/\delta_v$:

$$^0k^1_1 = {}^0k^2_2 = k, \quad k = \delta^2_v\left[1 - \frac{\cosh(y^*)}{\cosh(u)}\right], \tag{C1}$$

$^1\alpha^3 = -y_3$, other $^0k^i_j$ and $^1\alpha^i = 0$. After integrating on the gas layer $[-a+a]$ one obtains

$$^0K^1_1 = {}^0K^2_2 = \delta^2_v\left[1 - \frac{\tanh(u)}{u}\right], \quad \text{other } {}^0K^i_j = 0.$$

These solutions enable one to tackle the next problem $\{S^1_v\}$, which takes the simplified form:

$$S^1_v \begin{cases} i\omega\frac{\rho^e}{\mu}{}^1k^{mn}_i - {}^1k^{mn}_{i,33} - {}^2\alpha^{mn}_{,3} = {}^1\alpha^m\delta^n_i + 2{}^0k^m_{i,n}, \\ {}^1k^{mn}_{i,i} = -{}^0k^m_n + \langle^0k^m_n\rangle, \\ {}^1k^{mn}(\pm a) = 0, \quad \langle^2\alpha^{mn}\rangle = 0. \end{cases}$$

The resolution shows that the nonzero components read:

$$^1k^{11}_3 = {}^1k^{22}_3 = \frac{\delta^3_v}{\cosh(u)}\left[\sinh(y^*) - y^*\frac{\sinh(u)}{u}\right],$$

$$^1k^{13}_1 = {}^1k^{23}_2 = \delta^3_v\left[y^*\frac{\cosh(y^*)}{\cosh(u)} - u\frac{\sinh(y^*)}{\sinh(u)}\right],$$

$$^1k^{31}_1 = {}^1k^{32}_2 = \delta^3_v\left[u\frac{\sinh(y^*)}{\sinh(u)} - y^*\right],$$

$$^2\alpha^{11} = {}^2\alpha^{22} = \frac{\tanh(u)}{2u}\left[\frac{a^2}{3} - y^2_3\right],$$

$$^2\alpha^{33} = -\frac{1}{2}\left[\frac{a^2}{3} - y^2_3\right].$$

As the ^1k components are odd functions, they all are of zero mean value. Therefore,

$$^1K = 0.$$

The macroscopic tensors ^2K and ^2N are calculated with the expressions established in Appendix B that enable one to shun the explicit resolution of (S^2_v):

and, within the adiabatic assumption:

$$\gamma^2N^l_j = {}^0K^l_j + \langle^1\alpha^{j1}\alpha^l\rangle.$$

From the previous results, the expressions of the nonzero components become [k is defined in (C1)]:

$$^2K^{111}_1 = {}^2K^{222}_2 = \langle(k)^2\rangle + \langle^2\alpha^{11}k\rangle,$$

$$^2K^{122}_1 = {}^2K^{211}_2 = \langle(k)^2\rangle,$$

$$^2K^{112}_2 = {}^2K^{221}_1 = \langle^2\alpha^{11}k\rangle,$$

$$^2K^{133}_1 = {}^2K^{233}_2 = \langle(k)^2\rangle - 2\langle k_{,3}{}^1k^{13}_1\rangle,$$

$$^2K^{311}_3 = {}^2K^{322}_3 = \langle y_3{}^1k^{31}_1\rangle,$$

$$^2K^{131}_3 = {}^2K^{232}_3 = {}^2K^{313}_1 = {}^2K^{323}_2 = \langle y_3{}^1k^{13}_1\rangle,$$

$$^2K^{113}_3 = {}^2K^{223}_3 = {}^2K^{331}_1 = {}^2K^{332}_2 = \langle y_3{}^1k^{11}_3\rangle,$$

$$^2N^1_1 = {}^2N^2_2 = \frac{k}{\gamma}, \quad {}^2N^3_3 = \frac{\langle(^1\alpha^3)^2\rangle}{\gamma}.$$

Integration on the gas layer provides the expressions given in Sec. V.

Allard, J.-F. (**1993**). *Propagation of Sound in Porous Media* (Elsevier Applied Science, London and New York).

Attenborough, K. (**1983**). "Acoustical characteristics of rigid fibrous absorbents and granular media," J. Acoust. Soc. Am. **73**, 785–799.

Auriault, J. L. (**1980**). "Dynamic behaviour of a porous medium saturated by a Newtonian fluid," Int. J. Eng. Sci. **18**, 775–785.

Auriault, J. L. (**1991**). "Heterogeneous medium. Is an equivalent macroscopic description possible?," Int. J. Eng. Sci. **29**, 785–795.

Auriault, J. L., Borne, L., Chambon, R. (**1985**). "Dynamics of porous saturated media, checking of the generalized law of Darcy," J. Acoust. Soc. Am. **77**, 1641–1650.

Auriault, J. L., Geindreau, C., and Boutin, C. (**2005**). "Filtration law in

porous media with poor scale separation," Transp. Porous Media **60**, 89–108.

Bakhvalov, N., and Panasenko, G. (**1989**). *Averaging Processes in Periodic Media. Mathematical Problems in Mechanics of Composite Materials* (Kluwer Academic, Dordrecht).

Benssoussan, A., Lyons, J. L., and Papanicolaou, G. (**1978**). *Asymptotic Methods in Periodic Structures* (North-Holland, Amsterdam).

Biot, M. A. (**1956**). "Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178; 1956). "II. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–181.

Bond, L. J. (**1989**). "Numerical techniques and their use to study wave propagation and scattering - A review of elastic wave and ultrasonic non-destructive evaluation," *Proceedings of the IUTAM*, Boulder, CO.

Boutin, C., and Auriault, J. L. (**1990**). "Dynamic behaviour of porous media saturated by a viscoelastic fluid. Application to bituminous concrete," Int. J. Eng. Sci. **28**, 1157–1181.

Boutin, C., and Auriault, J. L. (**1993**). "Rayleigh scattering in elastic composite materials," Int. J. Eng. Sci. **31**, 1669–1689.

Boutin, C. (**1995**). "Microstructural influence on heat conduction," Int. J. Heat Mass Transfer **38**, 3181–3195.

Boutin, C. (**1996**). "Microstructural effect in elastic composite," Int. J. Solids Struct. **33**, 1023–1051.

Boutin, C., Royer, P., and Auriault, J. L. (**1998**). "Acoustic absorption of porous surfacing with dual porosity," Int. J. Solids Struct. **35**, 4709–4737.

Boutin, C. (**2000**). "Study of permeability by periodic and self consistent homogenisation," Eur. J. Mech. A/Solids **19**, 603–632.

Boutin, C., and Bazaille M. (**2005**). "Scattering of long acoustic wave in porous media," *Proceedings of the Third Biot Conference On Poromechanics*, Oklahoma, 24–27 May pp. 247–252.

Champoux, Y., and Allard, J. F. (**1991**). "Dynamic tortuosity and bulk modulus in air saturated porous media," J. Appl. Phys. **70**, 1975–1979.

Chen, W., and Fish, J. (**2001**). "A dispersive model for wave propagation in periodic heterogeneous media based on homogenization with multiple scales," J. Appl. Mech. **68**, 153–161.

Gambin, B., and Kröner, E. (**1989**). "High-order terms in the homogenized stress-strain relation in periodic elastic media," Phys. Status Solidi B **151**, 513–519.

Gubernatis, J. E., Domany, E., Krumhansl, J. A., and Hubernam, M. (**1977**). "The Born approximation in the theory of the scattering of elastc waves by flaws," J. Appl. Phys. **48**, 2812–2819.

Hirsekorn, S. (**1988**). "The scattering of ultrasonic waves by multiphase polycrystals," J. Acoust. Soc. Am. **83**, 1231-1242.

Ishimaru, A. (**1997**). *Wave Propagation and Scattering in Random Media* (IEEE, Oxford University Press Classic Reissue).

Kafesaki, M., and Economou, E. N. (**1999**). "Multiple-scattering theory for three-dimensional periodic acoustic composites," Phys. Rev. B **60**, 11993–12001.

Leclaire, P., Kelders, L., Lauriks, W., Glorieux, C., and Thoen, J. (**1996**), "Determination of the viscous characteristic length in air-filled porous materials by ultrasonic attenuation measurements," J. Acoust. Soc. Am. **99**, 1944–1948

Sanchez Palencia, E. (**1980**). *Nonhomogeneous Media and Vibration Theory*, Lectures Notes in Physics Vol. **127** (Springer, Berlin).

Sheng P. (**1995**). *Introduction to Wave Scattering, Localization and Mesoscopic Phenomena* (Academic Press, San Diego).

Stanke, F., and Kino, G. (**1984**). "A unified theory for elastic wave propagation in polycrystalline materials," J. Acoust. Soc. Am. **75**, 665–681.

Tournat, V., Pagneux, V., Lafarge, D., and Jaouen, L. (**2004**). "Multiple scattering of acoustic waves and porous absorbing media," Phys. Rev. E **70**, 026609:1-9.

Turbe, N. (**1982**). "Application of Bloch expansion to periodic elastic and viscoelastic media," J. Math. Meth. Appl. Sci. **4**, 433–449.

Zwikker, C., and Kosten, W. (**1949**). *Sound Absorbing Materials* (Elsevier, Amsterdam), 300p.

Reconstructing the adhesion stiffness distribution in a laminated elastic plate: Exact and approximate inverse scattering solutions

Ricardo Leiderman[a)]
Program of Mechanical Engineering, Federal University of Rio de Janeiro, Ilha do Fundo, P.B. 68509, Rio de Janeiro, RJ, Brazil 21945-970

Paul E. Barbone[b)]
Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

Arthur M. B. Braga[c)]
Department of Mechanical Engineering, Pontificia Universidade Catolica do Rio de Janeiro, Rua Marques de Sao Vicente, 225, Gavea, Rio de Janeiro, RJ, Brazil 22453-900

This paper formulates and solves a time harmonic inverse scattering problem to reconstruct the effective stiffness *distribution* of an adhesive bond in a layered elastic plate. The motivation is based on the assumption that localized adhesion flaws that diminish bond stiffness also tend to diminish bond strength. The formulation is based on the invariant imbedding method, applies to isotropic and anisotropic elastic layers, and is essentially that of identifying embedded acoustic sources in elastic layered structures. This paper presents two solutions for the inverse problem: the Born approximation and the exact solution. The example calculations compare the two solutions and show that when imperfections are too large in either magnitude or extent the accuracy of the Born approximation breaks down. The impact of noise and uncertainties in the background properties in the inversion is also investigated. A regularization strategy is introduced in the exact solution that controls solution sensitivity in regions with low signal to noise ratio.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2772212]

I. INTRODUCTION

We formulate and solve a time-harmonic inverse scattering problem to reconstruct the unknown *bond stiffness distribution* in a layered plate. We assume prior knowledge of the material properties of each layer, as well as the original (i.e., the undegraded) adhesive stiffness. We also assume knowledge of the incident field, which in our examples is taken to be a time-harmonic Gaussian beam in two dimensions. The "measured" signal is taken to be the surface displacement of the plate, as might be measured by optical techniques (cf. Ref. 1). The results show that, in principle, it is possible to measure the stiffness distribution over a large region with only a single insonification.

Related inverse problems were solved previously. The authors of Ref. 2 use surface wave dispersion measurements to back out bulk properties of an adhesive layer. In a series of papers,[3–5] Rokhlin and co-workers work up to the point of inferring not only the bulk properties of an embedded layer (e.g., the adhesive layer), but also the adhesion interface stiffness. They use data collected from transient pulses from both normal and oblique incidence, and assume transverse homogeneity of the sample. As such, the techniques from,

for example, Ref. 5 could serve in practice to provide the input "known" background properties for our inverse problem.

In the following we present two solutions of our inverse problem: one based on the Born approximation, and one based on the exact solution. The Born approximation is valid for weakly scattering inhomogeneities. As we show in our examples, the scattering from a bond defect can be weak either because the imperfection is small in magnitude, or because the imperfection is small in extent. In such cases the Born approximation closely matches the exact solution. If the defect is too large in one or the other, however, the accuracy of the Born approximation breaks down. In these cases, the exact solution can be used. It is scarcely more complicated than the Born approximation, and is always applicable.

The key to the solution involves treating the defect as a secondary source, and then seeking to evaluate the source strength. To do this, we work in the context of invariant embedding.[6]

As are most inverse problems, this is ill-conditioned. We avoid some degree of ill-conditioning by avoiding evanescent wave numbers. Even so, there are some regions of the plate where the incident field is vanishingly small. It is impossible to recover the correct stiffness in those regions, as

---
a)Electronic mail: leider@mecanica.ufrj.br
b)Electronic mail: barbone@bu.edu
c)Electronic mail: abraga@mec.puc-rio.br

any noise in the signal has a dramatic effect on the reconstructions there. We show how to control that high sensitivity with the addition of some regularization.

Section II contains a description of the mathematical model that we use. Following that, we state the inverse problem and present the solutions in four steps. Through several numerical examples, we demonstrate the inverse solutions, investigate the role of noise and regularization, and evaluate sensitivity of inversions to uncertainties in the plate's background properties.

II. FIELD EQUATIONS

The plate is considered to be a plane layered medium of infinite extent in two directions, but finite thickness, with an arbitrary number of layers. Each layer can be either isotropic or anisotropic. Between the adherends are adhesive layers. The regions above and below the plate are assumed to be occupied by an acoustic fluid. The case of a plate in vaccuum can be treated similarly.

We assume the z axis is normal to the layering, and designate the plane $z=0$ to be the "bottom" of the plate, and $z>0$ "upwards." In the fluid regions, $z<z_0=0$ and $z>z_N>0$, the fluid pressure satisfies

$$\nabla^2 p + k^2 p = 0, \tag{1}$$

$$p = p_{up} + p_{down}, \tag{2}$$

$$p_{up} = 0, \quad z < 0, \tag{3}$$

$$p_{down} = p_{inc} = \text{known}, \quad z > z_N. \tag{4}$$

In Eq. (2), we have decomposed the pressure field into up and downgoing components, respectively. The conditions (3) and (4) thus represent radiation conditions. In Eqs. (1)–(4) and in what follows, we assume and suppress a time dependence of $e^{-i\omega t}$.

The nth layer of the solid is supposed to occupy $z_{n-1} < z < z_n$, and have elasticity tensor C^n and density ρ^n. Therefore, the stress σ and displacement, u in the nth layer satisfy

$$\nabla \cdot \sigma + \rho^n \omega^2 u = 0, \quad z_{n-1} < z < z_n, \tag{5}$$

$$\sigma = C^n : \nabla u, \quad z_{n-1} < z < z_n. \tag{6}$$

We note that the nth layer may be either a structural layer or an adhesive layer.

Between each adhesive and adherend layer there is a thin interface of adhesion, which typically has thickness of microns, whose integrity controls the adhesion strength. It is modeled here by its effective stiffness in the form of spring constants and is depicted in Fig. 1. Such an approximation is useful when the layer (interface) of interest is much thinner than the inspecting wavelength.[7] The springs connect the adhesive to the adherend, and enforce continuity of *both* normal *and* shear traction and (approximately) transverse and normal displacement. It gives us the following boundary condition at the interface between the nth and $(n+1)$th layers:

$$K[u(z_n^+) - u(z_n^-)] = t(z_n^+), \tag{7}$$
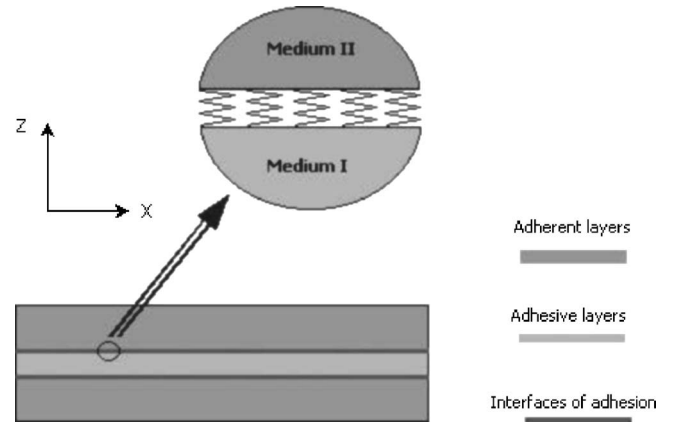
FIG. 1. Quasistatic approximation. When the inspecting wavelength is significantly larger than the interface thickness, the adhesion interface can be modeled as distributed transverse and normal springs.

$$t(z_n^+) = t(z_n^-). \tag{8}$$

Here K is a diagonal spring matrix, u is the displacement vector, and t is the traction vector acting on the xy plane. This "quasistatic" model is usually attributed to Baik and Thompson.[7] Expressions to calculate the entries in K in terms of the assumed thickness and elastic properties of the interfacial layer can be found in Ref. 8. Higher order extensions of similar models for thin layers have recently been described in Refs. 9–11. As u and t are vectors, Eq. (7) can model defects such as delamination and kissing bonds.

We assume that bond flaws are located at interfaces. Thus, imperfections in the adhesive bond between layers may be modeled by a reduction in the adhesive stiffness, K. Since imperfections tend to be localized in space, this implies that $K(x,y)$ must be allowed to depend upon position in the layer. A similar model for nonuniform interfacial stiffness can be found in Refs. 6 and 12. Therefore, we write:

$$K = K_0 + K_1(x,y). \tag{9}$$

Here K_0 is the background interfacial stiffness. It is constant and assumed to be known. The imperfection $K_1(x,y)$, on the other hand, is a function of the in plane position coordinates (x,y) and is the unknown in the inverse problem.

The interaction between imperfections and the incident field will give rise to the appearance of a scattered field. Thus, we rewrite Eqs. (7) and (8) as

$$[K_0 + K_1][\Delta u_{specular}(z_n) + \Delta u_{scatt}(z_n)]$$
$$= t_{specular}(z_n^+) + t_{scatt}(z_n^+), \tag{10}$$

$$t_{specular}(Z_n^+) + t_{scatt}(z_n^+) = t_{specular}(z_n^-) + t_{scatt}(z_n^-), \tag{11}$$

where

$$\Delta u(z_n) = u(z_n^+) - u(z_n^-). \tag{12}$$

The subscripts specular and scatt represent the specular field and the scattered field, respectively. By specular, we mean the field present in the absence of flaws, i.e., with $K_1=0$. By definition we can write:

$$K_0 \Delta u_{specular}(z_n) = t_{specular}(z_n^+), \tag{13}$$

$$t_{\text{specular}}(z_n^+) = t_{\text{specular}}(z_n^-). \tag{14}$$

Thus we can write the interface conditions for the scattered field as

$$K_0 \Delta u_{\text{scatt}}(z_n) + \varphi = t_{\text{scatt}}(z_n^+), \tag{15}$$

$$t_{\text{scatt}}(z_n^+) = t_{\text{scatt}}(z_n^-), \tag{16}$$

where

$$\varphi = K_1[\Delta u_{\text{specular}}(z_n) + \Delta u_{\text{scatt}}(z_n)]. \tag{17}$$

We can notice that Eq. (15) is similar to Eq. (13) except for the term φ. This term can be interpreted as a traction source acting at the interface. Therefore, a defective plate can be thought of as an ideal plate with secondary acoustic sources, i.e., those that are "activated" by the insonifying field. We took advantage of this to formulate the direct problem in Ref. 6 and will use the same strategy again for the present inverse problem. For this reason, both formulation and solution presented here are essentially that of identifying embedded acoustic sources in elastic layered structures.

III. INVERSE PROBLEM STATEMENT

In this section we consider the related inverse problem of reconstructing the effective stiffness distribution of internal adhesive bonds in a submerged layered elastic plate. As we are concerned primarily with adhesive bond degradation, we suppose that we know the material properties of each layer, and the undegraded adhesive stiffness K_0 of each bond. Similarly, we assume the incident sound field is known. Therefore, we can predict accurately the specular field, the field reflected by the plate in the absence of flaws. Any deviation between the "*measured*" reflected field and the idealized reflected field is thus attributed to the presence of flaws.

In the present work, we consider only one (anyone) of the adhesive interfaces to be defective. Further, we assume that we know *a priori* which interface is the defective one. It may be possible to determine which interface is the defective one by using multiple measurements, i.e., measurements for different frequencies or angles of incidence, but we restrict ourselves to the simplest case here. The problem where all interfaces are potentially defective is open at this point.

We summarize our problem as follows:
Given M^j, Z^j, K_0^j, for $j = 1, \ldots, N_{\text{layers}}$, and the acoustic fields measured at the top surface of the plate, find $K_1(x, y)$.

We note that the propagator matrices, M, and the impedance tensors, Z, define acoustically each layer. Expressions for them in terms of layer thickness and mechanical properties can be found in Ref. 6.

IV. SOLUTION OF THE INVERSE PROBLEM

For this solution we work in two dimensions and consider the xz plane to be the plane of propagation, as shown in Fig. 1. Some of the equations in the following indicate relations between Fourier transformed field variables. To that end, we use a single overbar to denote a single Fourier transform over the x coordinate.

Equation (17) relates the secondary source strength, φ, to the sought stiffness distribution and the interface displacement jump, $\Delta u(z_n)$. We aim to evaluate K_1 by independently evaluating φ and Δu, and then dividing as indicated by Eq. (17). The problem may be solved in four steps.

A. Step 1: Compute the scattered field at top surface

Given $u_{\text{total}}^{\text{top}}$ and $u_{\text{specular}}^{\text{top}}$, we may compute $u_{\text{scatt}}^{\text{top}}$ simply by

$$u_{\text{scatt}}^{\text{top}} = u_{\text{total}}^{\text{top}} - u_{\text{specular}}^{\text{top}}. \tag{18}$$

Here, $u_{\text{total}}^{\text{top}}$ is the total displacement field measured on the top surface of the defective plate. The field $u_{\text{specular}}^{\text{top}}$ is the hypothetical total displacement field that would be measured on the top of an identical plate with no flaws. For our calculations it was computed given the incident field and known plate properties; see the Appendix for details. In practice, one might consider measuring $u_{\text{specular}}^{\text{top}}$ on an undamaged plate or on an undamaged region of the same plate.

B. Step 2: Compute the secondary source strength

Our next step is to evaluate the source term, φ, from $u_{\text{scatt}}^{\text{top}}$. In Ref. 6 we have shown how to transport the effect of embedded sources to the boundary condition at the top of the plate. We summarize this in the Appendix. Thus we rearrange Eq. (A28):

$$\bar{\varphi} = -[W]^{-1}[G - Z_f n \otimes n]\bar{u}_{\text{scatt}}^{\text{top}}. \tag{19}$$

Equation (19) relates the embedded acoustic source to the resulting displacement field at the plate's top surface. Here Z_f is the fluid impedance, n is the unit vector in the z direction, G represents the surface impedance tensor looking into the top of the (unflawed) laminated plate, and W is the tensor which transports the effects of a traction source field from the defective interface up to the top surface. Expressions for Z_f, G, and W are derived in detail in Ref. 6 and are summarized in the appendix.

The bar over the field variables in Eq. (19) denotes a single Fourier transform over the x coordinate. It means that we transform $u_{\text{scatt}}^{\text{top}}$ evaluated via Eq. (18), find $\bar{\varphi}$ via Eq. (19), and inverse transform $\bar{\varphi}$ to find φ. φ is then used to reconstruct the defective bond stiffness, via its definition, as seen in the following.

C. Step 3: Backpropagate u_{specular} or u_{total}

At this point we must decide whether we wish to evaluate the exact solution or the Born approximation. To make clear the distinction between the two, we rewrite Eq. (17) as

$$\varphi = K_1 \Delta u_{\text{total}}(z_n) \tag{20}$$

$$\approx \varphi_B = K_1 \Delta u_{\text{specular}}(z_n). \tag{21}$$

To use the Born approximation, we evaluate $\Delta u_{\text{specular}}(z_n)$ and use Eq. (21) to evaluate K_1. To obtain the exact solution, we evaluate $\Delta u_{\text{total}}(z_n)$ and use Eq. (20) to evaluate K_1. The Born approximation has the advantage that $\Delta u_{\text{specular}}(z_n)$ may be computed once and for all, for any measured scattered field.

Whether we seek to evaluate $\Delta u_{\text{total}}(z_n)$ or $\Delta u_{\text{specular}}(z_n)$, the process is the same. To evaluate the displacement jump at the defective interface, $\Delta u(z_n)$, we must backpropagate the displacement field from the top of the plate down into the layered solid. At high frequencies, this process can be numerically unstable. Here we show how this backpropagation can be accomplished in an unconditionally stable way via invariant embedding.

We propagate the solution through each layer, one at a time, starting with the top layer. Clearly, then, the problem reduces to one of finding $\bar{u}(z_{j-1})$, the displacement field at $z=z_{j-1}$, given $\bar{u}(z_j)$, the total displacement field at $z=z_j$. To do that, we use Eqs. (A5)–(A7) evaluated at $z=z_j$ with the definition $\bar{u}_1+\bar{u}_2=\bar{u}$ to solve for $\bar{u}_2$:

$$\bar{u}_2(z_j) = [Z_2 - Z_1]^{-1}[[G_j - Z_1]\bar{u}(z_j) + W_j\bar{\varphi}]. \tag{22}$$

We can now use Eq. (A4) to propagate $\bar{u}_2(z_j)$ to the interface below, and use Eq. (A8) to find $\bar{u}_1(Z_{j-1})$:

$$\bar{u}_2(z_{j-1}) = M_2(z_{j-1} - z_j)\bar{u}_2(z_j), \tag{23}$$

$$\bar{u}_1(z_{j-1}) = R_{j-1}\bar{u}_2(z_{j-1}) + S_{j-1}\bar{\varphi}. \tag{24}$$

Once we have the solution at $\bar{u}(z_{j-1}^+)$ as given in Eqs. (23) and (24), we need to propagate the solution across the spring interface to evaluate $\bar{u}(z_{j-1}^-)$. This is most easily accomplished using continuity of traction, and the previously evaluated impedance tensors. In particular, using definition (A7) at both top and bottom of the interface gives

$$\bar{u}(z_{j-1}^-) = (G_{j-1}^-)^{-1}[G_{j-1}^+\bar{u}(z_{j-1}^+) + (W_{j-1}^+ - W_{j-1}^-)\bar{\varphi}]. \tag{25}$$

We emphasize again that $W^-=0$ for the defective interface. Also, we note that for the specular field calculations, $\bar{\varphi}=0$. This completes one full stage of the backward sweep. The total displacement field is advanced through each layer and each interface in turn, starting with the total solution at the top of the plate. Once the backward sweep is completed, the defect K_1 can be calculated by either the Born approximation or the exact solution.

D. Step 4: Evaluate defect magnitude

1. The Born approximation

To evaluate the adhesion stiffness K_1 by the Born approximation, we use

$$\varphi(x) \approx \varphi_B(x) = K_1(x)\Delta u_{\text{specular}}(x, z_n). \tag{26}$$

We note that the left-hand side of Eq. (26) is known, and that $\Delta u_{\text{specular}}(x, z_n)$ is known. Here we have explicitly indicated the x dependence of the dependent variables to emphasize that they vary with position. Since K_1 is diagonal, we get from Eq. (26) for each non-null displacement component, as *functions of x*:

$$K_1^{(1,1)}(x) \approx \frac{\varphi_B(x) \cdot e_x}{\Delta u_{\text{specular}}(x, z_n) \cdot e_x}, \tag{27}$$

$$K_1^{(2,2)}(x) \approx \frac{\varphi_B(x) \cdot e_y}{\Delta u_{\text{specular}}(x, z_n) \cdot e_y}, \tag{28}$$

$$K_1^{(3,3)}(x) \approx \frac{\varphi_B(x) \cdot e_z}{\Delta u_{\text{specular}}(x, z_n) \cdot e_z}, \tag{29}$$

where e_x, e_y, and e_z are the unit vectors in the x, y, and z directions, respectively.

2. The exact reconstruction

Here we use $\Delta \bar{u}_{\text{total}}(z_n)=\bar{u}(z_n^+)-\bar{u}(z_n^-)$. Specializing Eq. (25) to z_n and rearranging gives

$$\Delta \bar{u}_{\text{total}}(z_n) = \bar{u}(z_n^+) - (G_n^-)^{-1}[G_n^+\bar{u}(z_n^+) + W_n^+\bar{\varphi}]. \tag{30}$$

For the exact solution we work with φ given in Eq. (19) and $\Delta u_{\text{total}}(z_n)$ given in Eq. (30). Substituting these into Eq. (20) and recalling that K_1 is diagonal gives

$$K_1^{(1,1)}(x) = \frac{\varphi(x) \cdot e_x}{\Delta u_{\text{total}}(x, z_n) \cdot e_x}, \tag{31}$$

$$K_1^{(2,2)}(x) = \frac{\varphi(x) \cdot e_y}{\Delta u_{\text{total}}(x, z_n) \cdot e_y}, \tag{32}$$

$$K_1^{(3,3)}(x) = \frac{\varphi(x) \cdot e_z}{\Delta u_{\text{total}}(x, z_n) \cdot e_z}. \tag{33}$$

The solution is now complete.

E. Regularization and noise

As are many inverse problems, this is ill conditioned and sensitive to noise in the data. There are two places where that sensitivity can manifest itself. One is evident in Eq. (19), where we require the inverse of W. For some wave numbers and frequencies, however, W is singular or nearly singular. This is especially true for evanescent wave numbers, where the acoustic source energy decays exponentially with depth. Though W is singular, the vector on which W^{-1} acts is in its range, provided the data are perfect, and so the right-hand side of Eq. (19) remains theoretically well defined. When u_{scatt} is replaced by a measured value with some noise, however, it may be anticipated that some regularization may be required to ensure that Eq. (19) makes sense. In all the numerical simulations presented in Sec. V, however, we restricted our attention to nonevanescent wave numbers, and found no difficulty with the conditioning of W.

The second place that regularization is sometimes needed is when one of the denominators of Eqs. (27)–(29) or Eqs. (31)–(33) vanishes. This is as may be expected:. We can measure interface stiffness only in those areas where the interrogating wave amplitude is nonzero. Again, with noiseless data, the numerator can be expected to vanish at the same time as the denominator, making the fraction well defined in the limit. With noisy data, however, regularization is required as illustrated in the numerical examples in Sec. V.

V. SIMULATION RESULTS AND DISCUSSION

We show stiffness reconstructions of defective adhesion interfaces in this section. The chosen example corresponds to ultrasound inspection of a layered plate immersed in water. The modeled plate is composed of three layers: a copper

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Leiderman *et al.*: Inverse scattering from adhesive flaws   1909
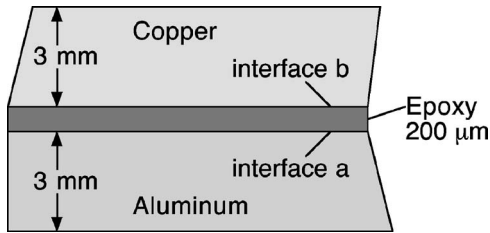
FIG. 2. The modeled plate. It has two adhesion interfaces, interfaces *a* and *b*.

layer with 3 mm thickness, an epoxy layer with 200 μm thickness acting as the adhesive layer, and an aluminum layer with 3 mm thickness. The plate is shown schematically in Fig. 2. It has, consequently, two adhesion interfaces: one between the copper and epoxy layers, interface *b*, and another between the epoxy and aluminum layers, interface *a*. The mechanical properties of the constituent materials are shown in Table I.

The adhesion interfaces are assumed to have nominal thicknesses of 3 μm each. When intact, they have the same mechanical properties as the epoxy. Accordingly, by the quasistatic model, these interfaces can then be represented by the following spring stiffness matrix:[8]

$$
K_0 = \begin{bmatrix} 0.4259 & 0 & 0 \\ 0 & 0.4259 & 0 \\ 0 & 0 & 1.8457 \end{bmatrix} 1 \times 10^{15} \text{ Pa/m}. \quad (34)
$$

All the modeled defects considered in the computations represent "kissing bonds" located at the interface *a*. Kissing bond is a region at the adhesion interface where there is a strong contact between the two media, but poor adhesion. The contact allows the bond to transmit normal traction and displacement, without the ability to transmit shear traction or in-plane displacement. In plane strain, a kissing bond can be modeled by diminishing the *xx* component (i.e., the in-plane component) of the original interfacial stiffness.

The incident field is taken to be a time-harmonic Gaussian beam with about 80 mm beam width, centered at $x=0$. It has a frequency of 4.9 MHz and angle of incidence of 3.82°. The frequency and angle of incidence were selected in order to maximize the scattered field, as described in Ref. 6. The incident wave excites a leaky Lamb mode in the plate. In Fig. 3 we show the normalized magnitude of the *x* component (the one that interacts with the defect) of the specular displacement jump at the defective interface, $|\Delta u_{x \text{ specular}}|$, for the region at the interface corresponding to -3 cm $>x$ >50 cm. Outside this region, the incident field and therefore the denominator of Eq. (27) or (31) is zero to working numerical precision.

Material	Density (kg/m³)	P-wave speed (m/s)	S-wave speed (m/s)
Aluminum	2700	6320	3130
Cooper	8930	4660	2660
Epoxy	1200	2150	1030
Water	1000	1480	0
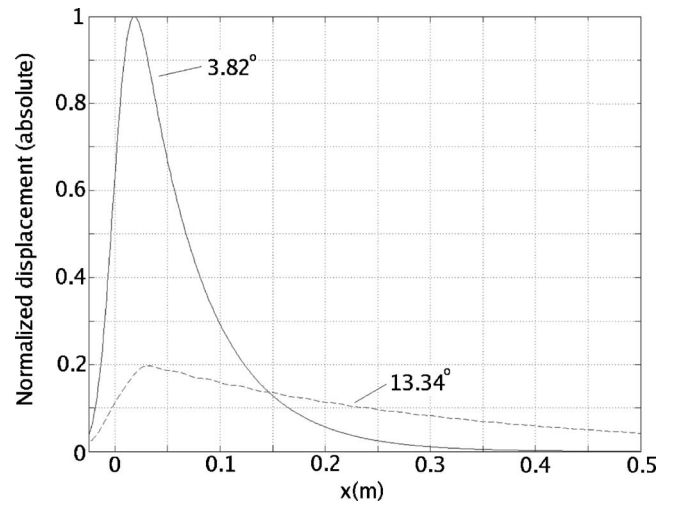
FIG. 3. The normalized magnitude of the *x* component of the specular displacement jump at the defective interface, $|\Delta u_{x \text{ specular}}|$, for 4.9 MHz and incidence at 3.82° (solid line) and at 13.34° (dashed line). We can see that the incidence corresponding to 13.34° excites a different leaky mode in the plate that has a lower amplitude in the vicinity of the incident wave beam, but a broader illuminated area.

merical precision. For this reason we restrict attention to the region -3 cm $>x>50$ cm (-3 cm $>x>25$ cm in Sec. V A).

A. Validation and reconstruction via the Born approximation

In Figs. 4 and 5 we show bond stiffness reconstructions via the exact solution and the Born approximation. The specular and scattered fields used here, as well as in all other performed computations, were computationally generated according to Ref. 6. The simulations were divided into two groups. In the first group the defect's length is the same but its magnitude varies. In the second group the defect's length varies while its magnitude remains the same. Results related
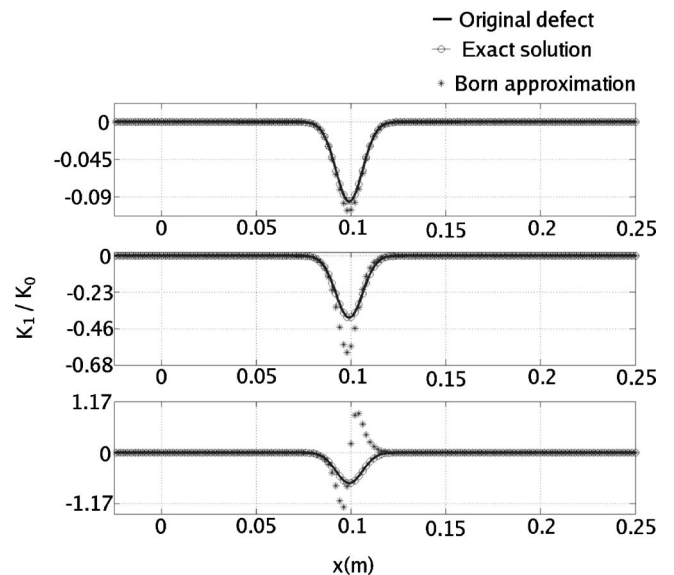
FIG. 4. Stiffness reconstructions via the exact solution and the Born approximation. Here, all defects have a Gaussian stiffness distribution extending about 40 mm in length. The magnitudes of the three defects are 10%, 40%, and 70% of the original bond stiffness, respectively.
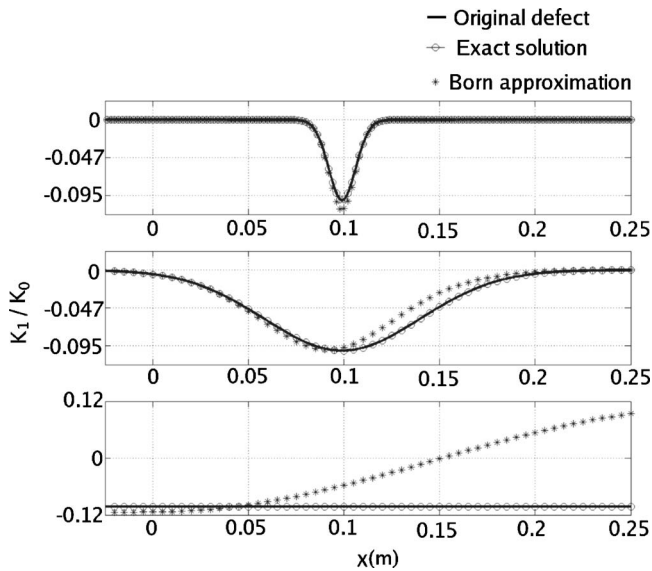
FIG. 5. Stiffness reconstructions via the exact solution and the Born approximation. Here, the magnitude of all defects is 10% of the original bond stiffness. They are all Gaussians in shape, with lengths 40, 250 mm, and ∞ (i.e., homogeneous), respectively.

to the first group of defects are presented in Fig. 4. In this group, all defects have Gaussian stiffness distribution extending about 40 mm in length. The magnitude of the three defects are 10%, 40%, and 70% of the original bond stiffness, respectively. Results related to the second group of defects are presented in Fig. 5. In this group, the magnitude of all defects was 10% of the original bond stiffness. They are all Gaussians in shape, with lengths 40, 250 mm, and ∞ (i.e., homogeneous), respectively.

In the figures we see that the exact solution reconstructs precisely all the modeled defects, validating our code. The Born approximation, however, does a poor job when the defect is large either in terms of extension or in terms of magnitude. The fact that the exact solution captures the homogeneous defect implies that the exact solution is insensitive to errors in K_0, the background adhesion stiffness.

B. The impact of noise in the inversion

To investigate the impact of noise in our inversion scheme, we synthetically introduced white noise in the scattered displacement field evaluated at the plate's top surface, and then reconstructed the bond stiffness via the exact solution.

1. Results for the original angle of incidence

In Fig. 6 we present inversion results for large, medium, and small signal to noise ratio (SNR), respectively, for the same defect reconstructed in the top plot of Fig. 4. Comparing Figs. 3 and 4 (top) shows the extent of the overlapping between the incident wavebeam and the defect. For large SNR, the noise has visible effect only for $x > 0.3$ m. In this region ($x > 0.3$ m), the incident field has magnitude comparable to the noise magnitude and, therefore, the *local SNR* is small. For medium SNR, the noise has visible effect for $x > 0.2$ m and for small *SNR*, the noise has visible effect in the
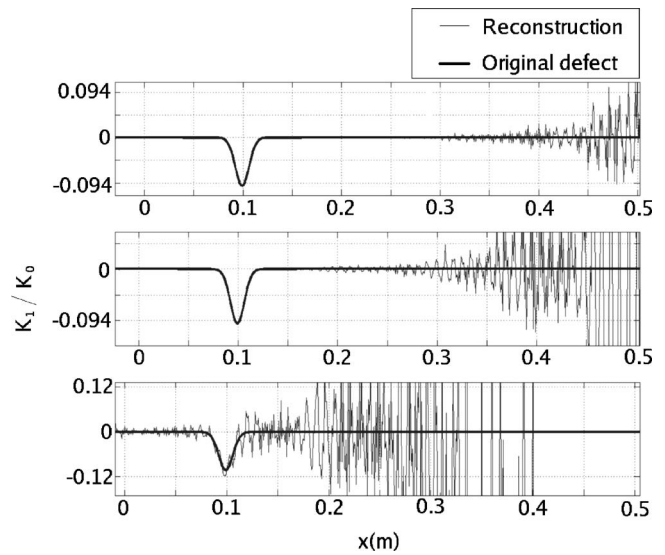
FIG. 6. Inversion results for large (≈ 4000), medium (≈ 800), and small (≈ 25) signal to noise ratio (SNR), respectively, for the defect shown in the top plot of Fig. 4. Comparing Figs. 3 and 4 (top) shows the extent of the overlap between the incident wave beam and the defect. For large SNR, the noise has a visible effect for $x > 0.3$ m, for medium SNR, the noise has visible effect for $x > 0.2$ m, and for small SNR, the noise has visible effect in the whole reconstruction region.

whole reconstruction extension. The plots show that the presence of noise is enhanced where the incident field magnitude is comparable to (or smaller than) the noise magnitude. As mentioned in Sec. IV E, it demonstrates that the inverse problem solution loses accuracy where the denominator of Eq. (31) approaches zero.

In Fig. 7 we investigate the effect of the degree of overlapping between the incident wavebeam and the defect, in the presence of noise. We computed the forward problem solution for the defect located at its original position, and shifted by 0.15 and 0.30 m to the right, and then computed the inverse problem solution for the three cases adding the same amount of noise (corresponding to a large SNR). In all
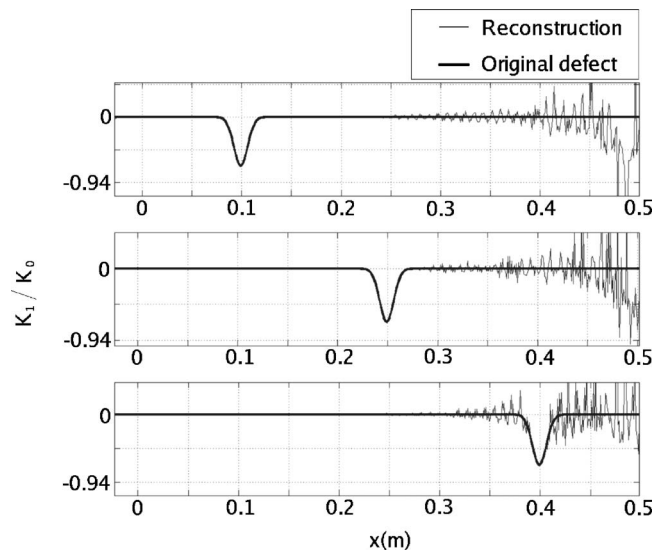
FIG. 7. Overlap between incident beam and defect. Defect is located at its original position, shifted by 0.15 m, and by 0.30 m to the right. The reconstructions are performed with noise at high SNR.
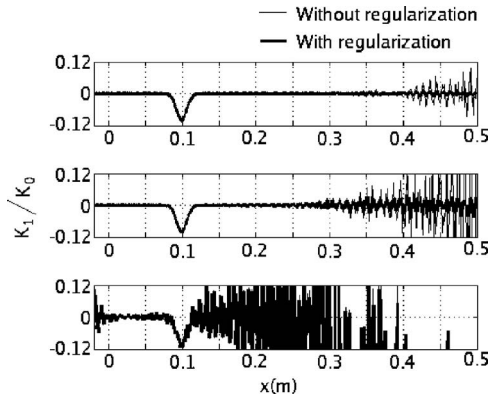
FIG. 8. Reconstructions for large (≈ 4000), medium (≈ 800), and small (≈ 25) SNR, respectively, the top, center, and bottom, with (thick solid line) and without (thin solid line) regularization. The regularization parameter was selected to improve the solution for a *large* SNR, i.e., $\epsilon/r_{\max} = 0.04 \ll 1$ has a small value.

FIG. 9. Same as Fig. 8 however the regularization parameter was adjusted to improve the solution for *small* SNR, i.e., $\epsilon/r_{\max} = 1 = O(1)$ has a large value.

plots, we see again that the noise has a greater impact where the incident field is weaker. The stiffness distribution is precisely reconstructed where the incident field is strong. In addition, a comparison between the plots shows that, to some extent, the stiffness distribution is better reconstructed at the region where the magnitude of the defect is large and thus the SNR is *locally* larger. This is illustrated in the plot at the bottom, where the defect was correctly reconstructed in the vicinity of $x = 40$ cm.

In practice, however, one would know *a priori* neither the defect location nor its magnitude. It means that the magnitude of the incident field is the primary factor that indicates where at the interface the noise compromises the solution, and determines, together with the SNR estimate, where the inverse solution is reliable.

The sensitivity of the solution to noise suggests that regularization is warranted. Here we introduce a regularization of the form:

$$K_1 = \frac{\varphi}{\Delta \boldsymbol{u}_{\text{reg}}}, \tag{35}$$

where

$$\Delta \boldsymbol{u}_{\text{reg}} = (r + \epsilon)\Delta \boldsymbol{u}/r, \tag{36}$$

$$r = \sqrt{\Delta \boldsymbol{u} \Delta \boldsymbol{u}^*} \tag{37}$$

and $\Delta \boldsymbol{u} = \Delta \boldsymbol{u}_{\text{specular}} + \Delta \boldsymbol{u}_{\text{scatt}}$ at the defective interface, ϵ is the regularization parameter proportional to the amount of noise present in the scattered field, and the asterisk means complex conjugate.

The regularization is designed to make the denominator of Eqs. (31)–(33) larger where it is too small and the inverse solution would otherwise be overly sensitive to noise. On one hand, to be effective, it has to be proportional to the amount of noise present in the measurements. On the other hand, it should be as small as possible to not "mask" the solution, i.e., to not diminish the defect's original magnitude in the reconstruction. The parameter ϵ in Eq. (36) is empiri-

cally adjusted based on these two principles. It is not always possible to find a suitable ϵ, especially when the SNR is small, as seen in the following.

In Fig. 8 we plot reconstructions for large, medium, and small SNR, respectively, at the top, center, and bottom of the figure, with (thick solid line) and without (thin solid line) regularization. The regularization parameter ϵ was adjusted to improve the solution for a *large* SNR, i.e., it has a small value. As expected, we see that the regularization significantly improved the solution in the plot shown at the top, improved to some extent the solution in the plot shown at the center, and had no effect in the plot shown at the bottom.

In Fig. 9 we plot the same, however the regularization parameter was adjusted to improve the solution for a *small* SNR, i.e., ϵ is large. In all the plots, we see that the regularization significantly improved the solution where the defect magnitude is small (practically zero), but the defect itself was under-reconstructed (compare to Fig. 4-top). In that sense, the comparison between Figs. 8 and 9 indicates that it is possible to correct for the noise in cases where the SNR is large, but the regularization tends to "mask" the defect in cases where the SNR is small.

Even a low value of regularization can mask the correct solution if the defect is located where the signal is weak, as shown in Fig. 10. In this figure we plot reconstructions for the defects shown in Fig. 7 (different degrees of overlapping between the incident field and the defect) for a large SNR (4000), with ϵ ($0.04 r_{\max}$) chosen accordingly and kept the
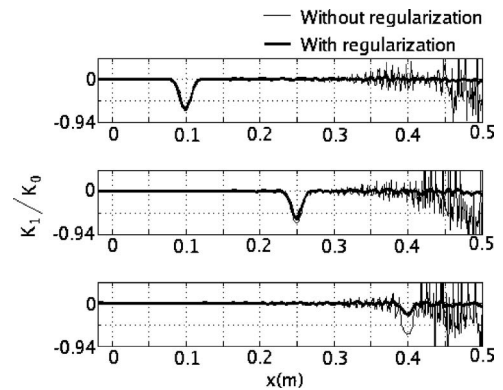
FIG. 10. Reconstruction for the defects shown in Fig. 7 for a large SNR (4000) and small $\epsilon (0.04 r_{\max})$.

Leiderman *et al.*: Inverse scattering from adhesive flaws
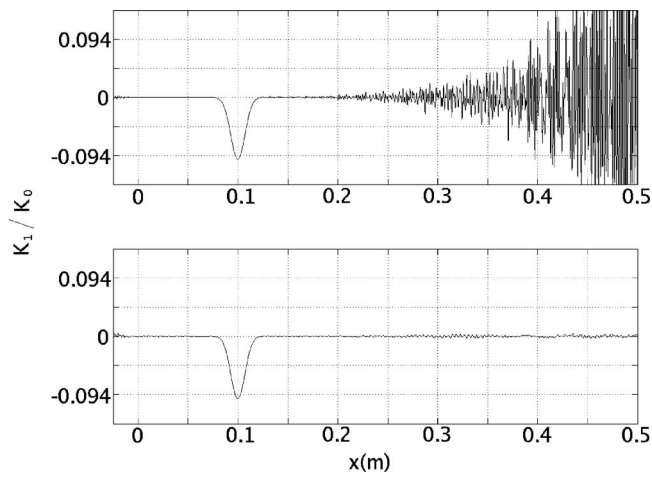
FIG. 11. Reconstructions for different angles of incidence. Low noise added to the data (SNR=4000), but no regularization used ($\epsilon=0$). Results related to 3.82° are ploted at the top, while the results related to 13.34° are plotted at the bottom. No regularization is needed for 13.34° incidence because the incident field is strong throughout the region.

same for all the three plots. For the plot shown at top we can see that the bond stiffness was well reconstructed in the whole interface extension. For the plot shown at the center, the bond stiffness was well reconstructed where the defect magnitude is negligible but the defect itself was slightly under-reconstructed. For the plot shown at the bottom, where there is the smallest degree of overlapping between the incident field and the defect, the defect was strongly under-reconstructed. It indicates that, beside masking the solution when the SNR is small, the regularization also tends to mask the solution where the incident field is weak.

2. Results for a different angle of incidence

Results corresponding to incidence at 3.82° and at 13.34° are compared here. In Fig. 3 we show the specular displacement field at the defective interface for both incident fields, normalized by the same factor. We can see that the incidence corresponding to 13.34° excites a different leaky mode in the plate that has a lower amplitude in the vicinity of the incident wave beam, indicating a weaker coupling between the incident field and kissing bonds, but a broader illuminated area. In fact, it illuminates a region at the interface larger than the one shown in the Fig. 3, indicating the potential to inspect a larger area. The weaker coupling between the incident field and the defect, however, makes the scattered signal weaker and more difficult to measure in practice, i.e., more susceptible to noise. We illustrate these effects in Fig. 11, where we show reconstructions for the defect shown in the top of Fig. 4 for large SNR. Results related to 3.82° are ploted at the top of the figure, while the results related to 13.34° are plotted at the bottom.

As before, we see again in the top of Fig. 11 that the introduction of noise has a stronger impact where the incident field is weaker. On the other hand, we show in the plot at the bottom that it has a negligible impact in the reconstruction corresponding to 13.34° angle of incidence. This noise
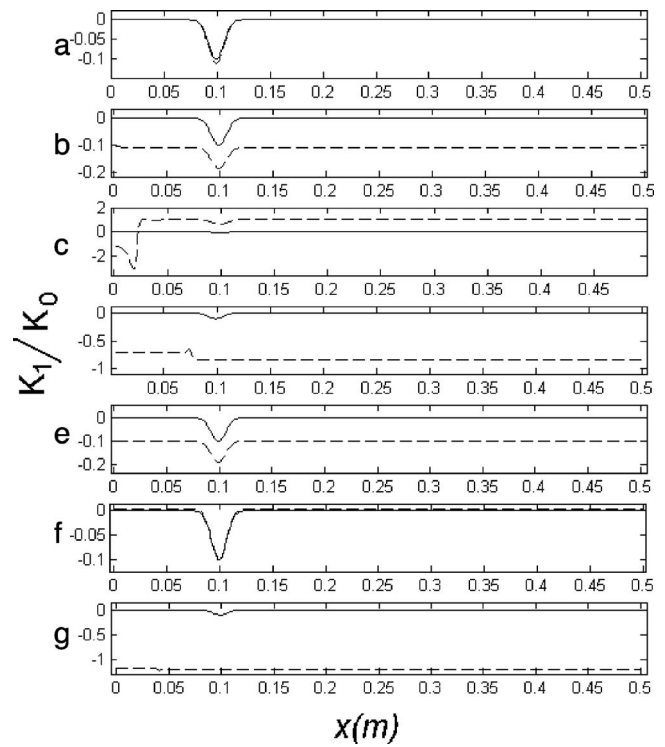
FIG. 12. Effects of miscalibration in back ground properties. Reconstructions obtained when performed with background properties that differ from those used in the forward solution are shown. (a) The defect was located at the interface b instead of interface a. (b)–(d) Five percent (under) miscalibration in the mass density, longitudinal wave speed, and shear wave speed of the copper layer, respectively. (e) Ten percent (under) miscalibration in the original bond stiffness (K_{0xx}). (f), (g) Ten percent (over) miscalibration in the longitudinal and shear wave speeds of the epoxy layer, respectively.

insensitivity reflects the fact that the incident field is relatively strong in the whole reconstruction extension (broader illuminated area) for 13.34°.

C. Impact of uncertainties in the plate's background properties

In the following, we investigate the impact of uncertainties in the plate's background properties on the inverse solution. For this discussion, we focus on the defect shown at the top of Fig. 4.

(a) *Uncertainty in position.* In order to evaluate the impact in uncertainty in the defect position, we solved the inverse problem assuming the defect was located at the interface b instead of interface a; see Fig. 2. The result, shown at the top of Fig. 12(a), shows that the impact of this assumption in the inversion is negligible with this choice of parameters.

(b) *Uncertainty in adherend material properties.* We next consider uncertainties in the background material properties. In the first example, we suppose that we attempted an inversion with the wrong mass density in the copper layer. To simulate this situation, we first solve the forward scattering problem with the mass density of the copper reduced by 5% from the value listed in Table I. Using this scattered field, we then reconstructed the adhesion stiffness using the original background properties (those listed in Table I). The resulting reconstructed adhesion stiffness (and the correct reconstruc-

tion) is shown in the second plot of Fig. 12 curve (b). Figure 12 shows a defect in the right place of the appropriate magnitude. It shows, erroneously, that the background value for the adhesion stiffness was overestimated by about 10%. Thus the background stiffness in the adhesion layer is to a certain extent compensating for the miscalibration of the mass density in the copper layer. We conclude from this example, as may have been expected, that uncertainties in the background properties can limit the quantitative accuracy of the reconstructions. Nevertheless, in this case we still clearly detect the presence of the defect.

The third and fourth plots of Figs. 12, curves (c) and (d), show, respectively, the corresponding results for uncertainties in the longitudinal and shear wave speeds in the copper layer. As before, the forward problem was solved with a reduction of 5% in the nominal value of the property under investigation. The adhesion stiffness was reconstructed using the nominal values of background properties. As before, the background value of the adhesion stiffness changes to partially compensate for the miscalibration in the background properties. These results are representative of the corresponding changes in the other layers.

(c) *Uncertainty in adhesion properties*. In the same fashion, we computed the scattered field considering a reduction of 10% in the interface a's original bond stiffness (K_{0xx}). We then solved the inverse problem assuming the original bond stiffness. The result, shown in the fifth plot of Fig. 12, curve (e), shows that this assumption has a relatively small impact on the inversion. In this case, the inversion correctly restores the background stiffness to its original value. Qualitatively speaking, curves (b) and (e) are very similar. In this case, the interpretation that the background adhesion stiffness was miscalibrated is correct; that same conclusion would be wrong if presented with curve (b).

(d) *Uncertainty in adhesive properties*. Of all the background properties, the adhesive stiffness is likely to be most variable. The last two curves in Fig. 12, curves (f) and (g), show the adhesion reconstructions with calibration errors of 10% in the longitudinal and shear wave speeds in the epoxy layer, respectively. We see that the reconstruction is relatively insensitive to the longitudinal wave speed in the epoxy, curve (f). The reconstruction is highly sensitive to the shear wave speed, curve (g). This sensitivity is to be expected, inasmuch as the insonifying mode was selected specifically to be sensitive to shear stress across the adhesion interface.

(e) *Discussion*. By the results plotted in Fig. 12 (and others not shown here), we can see that uncertainties in the material properties and geometeries of the plate constituents can dramatically impact the reconstructions. We note that the reconstructions are based upon an exact solution of the wave equations. Thus, the extent to which "the method" described here is sensitive to the problem parameters is the same as the extent to which the selected propagating mode is sensitive to the problem parameters.

The incident field used here is selected specifically to excite a particular leaky mode in the plate. We believe that the sensitivity to these properties that we see in the inversion arises through sensitivity of the excited leaky mode itself.

For example, the reconstruction is insensitive to the longitudinal wave speed in the epoxy, because the selected mode is primarily a shear mode in the epoxy. Different incident fields, therefore, will be sensitive to different plate properties to greater or lesser extents.

It is likely, for example, that a mode could be selected at a sufficiently low frequency such that the entire epoxy layer could be modeled accurately with the quasistatic approximation. In this case, the reconstruction method presented here would accurately reconstruct the spring properties that modeled even large variations in the epoxy properties. At such a low frequency, however, there would be some loss of spatial resolution without including evanescent modes in the analysis. Furthermore, at such low frequencies, it may be impossible to couple efficiently to the desired mode through water. A direct excitation of the plate may be desired in this case.

VI. SUMMARY AND CONCLUSIONS

We presented the exact and approximate solutions of an inverse scattering problem, corresponding to measuring the bond adhesion stiffness distribution in a laminated plate. Our solution gives the entire adhesion stiffness distribution along a line from a single insonification at a single frequency. At some point along the line of insonification, the incident wave amplitude diminishes to a negligible magnitude. The adhesion interface stiffness cannot be accurately measured beyond this point with the given insonification. We found a regularization strategy that could control the sensitivity of the reconstructed defect in this region.

Our inverse problem solution depends upon knowing *a priori* the background properties of the plate, and is based on assuming that exactly one interface is defective. If N interfaces were defective, we would need to evaluate strengths of N secondary sources. This requires at least N different measurements. These could come from N different incident fields, at say different frequencies or angles of incidence. A transient pulse at a single angle of incidence contains many frequencies, and may contain sufficient information to determine the stiffnesses of many interfaces. In any case, different interfaces would receive different incident fields. Thus the accuracy of their stiffness reconstructions would be different.

In the examples treated here, the solutions are highly sensitive to some of the background properties. This implies that if this inverse problem solution is to be usable in practice, with this choice of parameters, some robust technique(s) will be needed to evaluate the bulk properties independently. It is likely that different choices of incident fields could be selected that would be more or less sensitive to specific properties in a given problem. Some additional robustness of the inverse solution may come from the use of additional data to overconstrain the reconstruction.

ACKNOWLEDGMENTS

Leiderman *et al.*: Inverse scattering from adhesive flaws

APPENDIX: PROPAGATOR AND IMPEDANCE MATRICES

In the invariant embedding method, we decompose the total displacement and traction fields into upgoing fields, $\bar{u}_1(z)$, and downgoing field, $\bar{u}_2(z)$, according to their wave numbers in the vertical (z) direction, viz.,

$$\bar{u}(z) = \bar{u}_1(z) + \bar{u}_2(z), \tag{A1}$$

$$\bar{t}(z) = \bar{t}_1(z) + \bar{t}_2(z). \tag{A2}$$

From the exact solution of the elastodynamic equations of motion in homogeneous media, we may define the operators M_1, M_2, Z_1, and Z_2, such that

$$\bar{u}_1(z) = M_1(z - z_1)\bar{u}_1(z_1), \tag{A3}$$

$$\bar{u}_2(z) = M_2(z - z_1)\bar{u}_2(z_1), \tag{A4}$$

$$\bar{t}_1(z) = -i\omega Z_1\bar{u}_1(z), \tag{A5}$$

$$\bar{t}_2(z) = -i\omega Z_2\bar{u}_2(z). \tag{A6}$$

The propagator matrices M_j contain only propagating or decaying exponentials, as long as $z - z_1$ is positive in Eq. (A3) and negative in Eq. (A4). It helps to keep the method numerically stable. The *local impedance tensors*, Z_j, relate up and downgoing tractions to the corresponding displacement fields. Explicit expressions for these quantities can be found in Ref. 6.

In what follows, we show how to compute recursively the surface impedance and traction fields acting on the top surface of a layered plate, in terms of the corresponding values on the bottom of the plate and the properties of the plate itself. To that end, we begin by showing the computations for an elastic layer resting on a (layered) substrate.

1. Elastic layer

Here we consider an elastic layer with thickness h resting on a substrate. The bottom of the layer is located at $z = z_0$. The top of the layer is $z = z_1 = z_0 + h$. The substrate upon which the layer rests has a known surface impedance, G_0, and traction source field, $\bar{\varphi}$. At $z = z_0$, we write the traction vector, $\bar{\bar{t}}$, as

$$\bar{t}(z_0) = -i\omega[G_0\bar{u}(z_0) + W_0\bar{\varphi}], \tag{A7}$$

where W_0 is the matrix that transports the effects of an internal acoustical source to the bottom of the elastic layer. *For the present case, where only one source is considered, $W_0 = 0$ for all interfaces situated below the defective one.*

Now, we define the displacement field at $z = z_0$ as

$$\bar{u}(z_0) = \bar{u}_2(z_0) + R_0\bar{u}_2(z_0) + [Z_1 - G_0]^{-1}W_0\bar{\varphi}. \tag{A8}$$

Here, R_0 is the reflection matrix. By considering the continuity of displacement and traction, from Eq. (A8) in Eqs. (A5) and (A6) [we note that the last two terms of the right-hand side of Eq. (A8) are upgoing fields] and from definition (A2), we have

$$R_0 = -[Z_1 - G_0]^{-1}[Z_2 - G_0]. \tag{A9}$$

Finally, by forward- and backpropagating the up and downgoing waves (A8) up to $z = z_1$, we can write the traction vector as

$$\bar{t}(z_1) = -i\omega[G_1\bar{u}(z_1) + W_1\bar{\varphi}], \tag{A10}$$

where

$$G_1 = [Z_2 + Z_1R_0(z_1)][I + R_0(z_1)]^{-1}, \tag{A11}$$

$$W_1 = [Z_1 - G_1]M_1(h)[Z_1 - G_0]^{-1}W_0, \tag{A12}$$

$$R_0(z_1) = M_1(h)R_0M_2(-h). \tag{A13}$$

2. Adhesion interfaces

As before, here we consider an adhesion interface in contact with an elastic layer. The layer may be an adherend or may be a layer of the adhesive itself, depending on the frequency range of interest. The adhesion interface is modeled as a layer of springs, as described earlier. At this point it is useful to distinguish between defective and nondefective interfaces. For defective interfaces, the interface condition is

$$K_0[\bar{u}^+ - \bar{u}^-] + \bar{\varphi} = t^+ = t^-. \tag{A14}$$

The superscript $+$ indicates the value of the field variables above the interface, while the superscript $-$ indicates those below. We recall that K_0 represents the undegraded, or background, interface stiffness.

The layered substrate upon which the interface spring layer rests has a known surface impedance, such as calculated in Eq. (A11). The impedance boundary condition is given by

$$\bar{t}^- = -i\omega G^-\bar{u}^-. \tag{A15}$$

It is implicit in the condition above that $W^- = 0$.

To find the equivalent boundary condition at the top of the adhesive interface, we solve Eq. (A14) for $\bar{u}^-$ and substitute into Eq. (A15) to find

$$\bar{t}^+ = -i\omega[G^+\bar{u}^+ + W^+\bar{\varphi}], \tag{A16}$$

$$G^+ = [I + i\omega G^-K_0^{-1}]^{-1}G^-, \tag{A17}$$

$$W^+ = G^+K_0^{-1}. \tag{A18}$$

Note that in Eq. (A16) we introduced W^+. In fact, in the recursive computations, this is the first location (the top of the defective adhesive interface) where the tensor W is calculated.

For nondefective interfaces located above the defective one, Eqs. (A16) and (A17) hold. Equation (A18) changes to

$$W^+ = [I + i\omega G^+K_0^{-1}]W^-. \tag{A19}$$

The quantities calculated by Eqs. (A16)–(A19) can be substituted into Eq. (A7) to move forward with the computations for an elastic layer resting on the interface spring layer. Therefore, Eqs. (A7)-(A19) give all the ingredients needed to compute recursively the surface impedance and traction fields acting on the top surface of a layered plate, in terms of the corresponding values on the bottom of the plate

and the properties of the plate itself.

3. Reflection and radiation from top surface

Once we can compute $\boldsymbol{G}$ and $\boldsymbol{W}$, the impedance and the source transport matrix, we can formulate the reflection and radiation problem at the top surface of the plate. The boundary conditions in terms of the fluid pressure $\wp$ and fluid normal displacement ω_f at the top interface are

$$\bar{\boldsymbol{t}} = -\wp \, \boldsymbol{n}, \tag{A20}$$

$$\boldsymbol{n} \cdot \bar{\boldsymbol{u}} = w_f. \tag{A21}$$

These conditions enforce continuity of traction and normal displacement at the fluid/solid interface. Of course, we also have our impedance condition on the top surface of the solid:

$$\bar{\boldsymbol{t}} = -i\omega[\boldsymbol{G}\bar{\boldsymbol{u}} + \boldsymbol{W}\bar{\boldsymbol{\varphi}}]. \tag{A22}$$

Substituting Eq. (A20) into Eq. (A22) and solving for $\bar{\boldsymbol{u}}$ gives

$$\bar{\boldsymbol{u}} = \boldsymbol{G}^{-1}[(i\omega)^{-1}\wp \, \boldsymbol{n} - \boldsymbol{W}\bar{\boldsymbol{\varphi}}]. \tag{A23}$$

Equation (A23) can be used later to give the displacement field at the top of the plate, once the pressure is known.

As in every other layer, we decompose the pressure in the fluid into downgoing (incident) and upgoing (radiated) field,

$$\wp = \wp^{\mathrm{inc}} + \wp^{\mathrm{rad}}. \tag{A24}$$

Following steps analogous to those that lead to Eq. (A10) leads us to

$$\wp^{\mathrm{rad}} = r\wp^{\mathrm{inc}} + i\omega \frac{Z_s Z_f}{Z_s + Z_f} \boldsymbol{n} \cdot \boldsymbol{G}^{-1}\boldsymbol{W}\bar{\boldsymbol{\varphi}}, \tag{A25}$$

$$r = \frac{Z_s - Z_f}{Z_s + Z_f}, \tag{A26}$$

$$Z_s = [\boldsymbol{n} \cdot \boldsymbol{G}^{-1} \cdot \boldsymbol{n}]^{-1}, \tag{A27}$$

where r is the scalar representing the reflection coefficient at the plate's upper surface that relates the incident pressure field to its specular reflection. The right-hand side of Eq.

(A25) is known. At this point, the solution for the direct problem is formally known.

We recall that Eqs. (A23)-(A27) can be used to determine the specular displacement field from the pressure field by making $\bar{\boldsymbol{\varphi}}=\boldsymbol{0}$. Likewise, recognizing that in the fluid, where only P-waves propagate, the upgoing pressure is related to the upgoing vertical (z) displacement component, $\bar{w}_{f2}$, by $-\wp_1(z)=i\omega Z_f \bar{w}_{f1}(z)$, and making $\wp^{\mathrm{inc}}=0$, Eq. (A23) gives

$$Z_f \boldsymbol{n} \otimes \boldsymbol{n}\bar{\boldsymbol{u}}_{\mathrm{scatt}} = [\boldsymbol{G}\bar{\boldsymbol{u}}_{\mathrm{scatt}} + \boldsymbol{W}\bar{\boldsymbol{\varphi}}]. \tag{A28}$$

At this point, the solution for the inverse problem of identifying an embedded acoustic source is formally known.

[1] J. W. Wagner, "Optical detection of ultrasound," in *Physical Acoustics* (Academic, New York, 1990) Vol. **XIX**, pp. 210–266.

[2] T.-T. Wu and Y.-H. Liu "Inverse determinations of thickness and elastic properties of a bonding layer using laser-generated surface waves," Ultrasonics **37**, 23–30 (1999).

[3] A. I. Lavrentyev and S. I. Rokhlin, "Ultrasonic spectroscopy of imperfect contact interfaces between a layer and two solids." J. Acoust. Soc. Am. **103**, 657–664 (1998).

[4] 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).

[5] A. Baltazar, L. Wang, B. Xie, and S. I. Rokhlin, "Inverse ultrasonic determination of imperfect interfaces and bulk properties of a layer between two solids," J. Acoust. Soc. Am. **114**, 1424–1434 (2003).

[6] R. Leiderman, P. E. Barbone, and A. M. B. Braga, "Scatting of ultrasonic waves by defective adhesion interfaces in submerged laminated plates," J. Acoust. Soc. Am. **118**, 2154–2166 (2005).

[7] J. M. Baik and R. B. Thompson, "Ultrasonic scattering from imperfect interfaces: A quasi-static model," J. Nondestruct. Eval. **14**, 177–196 (1984).

[8] S. I. Rokhlin and W. Huang, "Ultrasonic wave interaction with a thin layer between two anisotropic solids," J. Acoust. Soc. Am. **92**, 1729–1742 (1992).

[9] D. D. Zakharov, "High order approximate low frequency theory of elastic anisotropic lining and coating," J. Acoust. Soc. Am. **119**, 1961–1970 (2006).

[10] L. Wang and S. I. Rokhlin, "Modeling of wave propagation in layered piezoelectric media by a recursive asymptotic method," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 1060–1071 (2004).

[11] L. Wang and S. I. Rokhlin, "Recursive geometric integrators for wave propogation in a functionally-graded multilayered elastic medium," J. Mech. Phys. Solids **52**, 2473–2506 (2004).

[12] S. Nakagaua, K. T. Nihei, and L. R. Myer, "Plane wave solution for elastic wave scattering by heterogeneous fracture," J. Acoust. Soc. Am. **115**, 2761–2772 (2004).

Nonlinear shear wave interaction in soft solids

Xavier Jacob, Stefan Catheline, Jean-Luc Gennisson,[a] Christophe Barrière,
Daniel Royer, and Mathias Fink
*Laboratoire Ondes et Acoustique, ESPCI-CNRS UMR 7587-INSERM-Universite Paris VII, 10 rue Vauquelin
75231 Paris Cedex 05, France*

This paper describes nonlinear shear wave experiments conducted in soft solids with transient elastography technique. The nonlinear solutions that theoretically account for plane and nonplane shear wave propagation are compared with experimental results. It is observed that the cubic nonlinearity implied in high amplitude transverse waves at $f_0 = 100$ Hz results in the generation of odd harmonics $3f_0$, $5f_0$. In the case of the nonlinear interaction between two transverse waves at frequencies f_1 and f_2, the resulting harmonics are $f_i \pm 2f_j (i,j=1,2)$. Experimental data are compared to numerical solutions of the modified Burgers equation, allowing an estimation of the nonlinear parameter relative to shear waves. The definition of this combination of elastic moduli (up to fourth order) can be obtained using an energy development adapted to soft solid. In the more complex situation of nonplane shear waves, the quadratic nonlinearity gives rise to more usual harmonics, at sum and difference frequencies, $f_i \pm f_j$. All components of the field have to be taken into account. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2775871]

I. INTRODUCTION

The characterization of soft matter mechanical properties is of interest for medical applications. For example, shear elasticity is a pertinent indicator for tissue degradation diagnosis. Research efforts also lead to evaluation of anisotropy,[1] viscosity,[2] and nonlinearity.[3] Linear and nonlinear elastic modulus of soft materials can be obtained by acoustoelasticity (measurements of speed variations under static strains) or by dynamic experiments (imaging low frequency shear wave propagation). The second way goes through solving an inverse problem in the linear case or fitting experimental and theoretical results in the nonlinear one.

Among these methods, finite amplitude experiments are interesting for several reasons. First, it gives practical solutions to measure the elastic nonlinearity and second, from a physical point of view, it is the first direct experimental observation of a shock transverse wave.[4] Indeed, in most "hard solids" such as metals or crystals, the particle velocity induced by acoustic source is very small compared to the wave speeds. Thus the small Mach numbers reached in experiments do not allow one to observe the tiny third-order nonlinear effects induced by transverse waves (propagation in highly inhomogeneous material such as rocks is not tackled in this paper, but can also lead to odd harmonic generation through the nonlinearity of the microstructure). To our knowledge, the same reasons avoided any experimental observation of nonlinear interaction between two transverse waves. Conversely, mechanical vibrators used in transient elastography experiments on soft solids allow one to reach Mach numbers close to unity.

This paper first recalls nonlinear equations obtained with quadratic and cubic approximation describing, respectively, longitudinal and transverse plane wave nonlinear propagation. Then, the specific model for elastic response of soft solids like agar-gelatin phantoms is examined. Considering the quasi-incompressibility of the medium, valid for shear solicitations, allows one to express the elastic energy in a simplified form.[5] Such a development defines elastic moduli relative to nonlinear shear elasticity, and can be used to established, in a simple way, a Burgers like equation, referred to as modified Burgers equation, relative to shear waves.

In the second part, elastographic experiments, which allow one to quantify these properties, are presented. Experimental results, obtained with harmonic generation and nonlinear interaction of quasi-plane shear waves, are compared to numerical solutions of the modified Burgers equation. Cubic nonlinear effects (odd harmonics) are clearly revealed through wave forms and spectra, and the nonlinear coefficient is estimated. Nonplane shear waves are also studied from the experimental point of view. The effect of the quadratic nonlinearity (even harmonics) also appears in this case. The link between quadratic nonlinear effect and diffraction can be understood through a Gaussian beam model.

II. MODEL EQUATIONS

A. Quadratic and cubic approximations

Elastic (or acoustic) nonlinearity is classically introduced by the mean of series expansions of the stress-strain relation (or state equation), taking into account the large amplitude of the particle velocity or of the strain deformation. In order to obtain a nonlinear equation that takes into account second- and third-order terms (of displacement derivatives), the elastic energy has to be developed to fourth order in strains.[6,7] The first intervention of nonlinearity is the complete definition of the strain itself:

---

[a] Electronic mail: jl.gennisson@espci.fr

$$S_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial a_j} + \frac{\partial u_j}{\partial a_i} + \frac{\partial u_k}{\partial a_i}\frac{\partial u_k}{\partial a_j}\right), \tag{1}$$

where the displacement $u_i = x_i - a_i$, is the difference between the current position x_i of the particle and its position at rest a_i, i.e., the Lagrangian coordinates.

The state equation (stress-strain relation) obtained by using such an expansion [Eq. (1)] leads to a wave equation with quadratic and cubic nonlinear terms. At the fourth order of the strain tensor, the elastic energy density takes the general form:

$$W = \frac{1}{2!}C^s_{ijkl}S_{ij}S_{kl} + \frac{1}{3!}C^s_{ijklmn}S_{ij}S_{kl}S_{mn}$$
$$+ \frac{1}{4!}C^s_{ijklmnpq}S_{ij}S_{kl}S_{mn}S_{pq}, \tag{2}$$

which gives a stress tensor at third order in displacement gradient $\partial u_i / \partial a_j$.

In the isotropic case, the energy can be expressed as a function of the three scalar invariants of $\bar{S}$ (the invariants defined by Landau, usually denoted I_1, I_2, and I_3, are the respective traces of $\bar{S}$, $\bar{S}^2$, and $\bar{S}^3$):[7]

$$W = \mu I_2 + \frac{\lambda}{2}I_1^2 + \frac{A}{3}I_3 + BI_1I_2 + \frac{C}{3}I_1^3 + EI_1I_3 + FI_1^2I_2 + GI_2^2$$
$$+ HI_1^4. \tag{3}$$

This form takes into account both compression and shear strains, with a complete set of elastic constants or moduli: two at the second order (linear elastic approximation expressed with the Lamé coefficients λ and μ), three at the third order A, B, C, and four at the fourth order E, F, G, H.

Then the strain energy can be expressed as a function of the principal invariants, I_S, II_S, and III_S defined by the Cayley-Hamilton theorem, i.e., the coefficients of the characteristic equation of $\bar{S}$:

$$W_s = \frac{1}{2}(\lambda + 2\mu)I_s^2 + 2\mu II_s + \frac{1}{3}(A + 3B + C)I_s^3 + (A$$
$$+ 2B)I_sII_s + AIII_s + (E + F + G + H)I_s^4 + 3EI_sIII_s$$
$$+ (3E + 2F + 4G)I_s^2II_s + 4GII_s^2. \tag{4}$$

In this development, it is possible to distinguish, at third order, invariants and moduli relative to pure mode propagation. The first line of Eq. (4) gives the velocities of pure longitudinal ($\sqrt{(\lambda+2\mu)/\rho_0}$, L waves) and shear waves ($\sqrt{\mu/\rho_0}$, T waves) in the linear approximation. The I_S^3 term [of coefficient $(A+3B+C)$] corresponds to the quadratic nonlinearity of L waves, the combination I_sII_s is linked to the coupling between T waves and L waves and III_s is related to the T waves.

In Lagrangian coordinates, the stress tensor can be written as[6]

TABLE I. Relations between the Landau moduli and the elastic constants for an isotropic solid (two first columns). For a purely compressible media, the third column gives the relation between these elastic constants and the coefficient defined for a perfect fluid (with a pressure-density state equation).

Landau	Elastic constant	Fluid
A	$4C_{456}$	0
B	$C_{144} = C_{166} - 2C_{456}$	$-A_l$
$2C$	$C_{123} = C_{111} - 6C_{166} + 4C_{456}$	$A_l - B_l$
$2(A+3B+C)$	C_{111}	$-(5A_l + B_l)$
$A+2B$	$2C_{166}$	$-2A_l$
$2B+2C$	$C_{112} = C_{111} - 4C_{166}$	$-(A_l + B_l)$

$$T_{ik} = \frac{\partial W}{\partial\left(\dfrac{\partial u_i}{\partial a_k}\right)}. \tag{5}$$

The wave equation follows from the momentum conservation law:

$$\rho_0 \frac{d^2 u_i}{dt^2} = \frac{\partial T_{ik}}{\partial a_k}. \tag{6}$$

In Lagrangian frame, all nonlinear terms come from the right-hand side of Eq. (6).

First, we recall principal features of the quadratic approximation for plane and nonplane waves. Considering a wave propagating along a_1 leads to the following system of equations:

$$\frac{\partial^2 u_1}{\partial a_1^2} - \frac{1}{V_L^2}\frac{\partial^2 u_1}{\partial t^2} = 2\beta_L\frac{\partial u_1}{\partial a_1}\frac{\partial^2 u_1}{\partial a_1^2} + \beta_{LT}\left(\frac{\partial u_2}{\partial a_1}\frac{\partial^2 u_2}{\partial a_1^2}\right.$$
$$\left. + \frac{\partial u_3}{\partial a_1}\frac{\partial^2 u_3}{\partial a_1^2}\right), \tag{7}$$

$$\frac{\partial^2 u_i}{\partial a_1^2} - \frac{1}{V_{T^2}}\frac{\partial^2 u_i}{\partial t^2} = \beta_{TL}\left(\frac{\partial u_1}{\partial a_1}\frac{\partial^2 u_i}{\partial a_1^2} + \frac{\partial u_i}{\partial a_1}\frac{\partial^2 u_1}{\partial a_1^2}\right), \tag{8}$$

with $i = 2, 3,$

where

$$\beta_L = -\frac{3(\lambda + 2\mu) + 2(A + 3B + C)}{2(\lambda + 2\mu)},$$

$$\beta_{TL} = -\frac{2(\lambda + 2\mu) + (A + 2B)}{2\mu} = \frac{\mu}{\lambda + 2\mu}\beta_{LT}. \tag{9}$$

In order to obtain Eq. (9) as a function of the elastic moduli, Table I recalls the correspondence between elastic moduli and the coefficient defined for a fluid up to the third order. Following Gol'dberg,[8] in the case of L plane waves, terms weighted by β_{LT} and β_{TL} vanish. Then Eq. (7) reduces to the quadratic nonlinear equation of L waves (quasiequivalent to the one obtained for fluids). In the case of T plane waves, Eq. (8) reduces to the linear wave equation for transverse waves ($\partial u_1/\partial a_1 = 0$). However the condition ($\partial u_3/\partial a_1$) $\times(\partial^2 u_3/\partial a_1^2) = 0$ imposes the longitudinal component not to be uniformly zero. Then, from quadratic approximation, the

propagation of an initially transverse polarization can produce a longitudinal component at double frequency. This effect is not cumulative (longitudinal and transverse waves propagate at different velocities, so phase matching conditions are not fulfilled) and, in the case of a gel, this quadratic coupling does not results in significant amplitudes because of the large difference between compression and shear moduli. (For nonunidirectional propagation, the different interaction cases between L and/or T plane waves have been examined,[9,18] considering as well intersecting waves. Depending on polarization, elastic moduli, and angles of incidence, cumulative interactions are possible, even at second order.)

Finally, the quadratic approximation does not allow one to deal with a nonlinear effect relative to plane transverse waves. The model presented in Sec. II B deals with the cubic nonlinearity but neglects the longitudinal component and the second-order transverse component.

B. Cubic approximation and soft media elasticity

It has been pointed out in previous papers, from theoretical[5,10] and experimental[11] points of view, that soft solids can be considered as quasi-incompressible. The physical interpretation of shear moduli can be attributed to μ and A, because both are zero for a perfect fluid. This can be obtained by comparing the energy development, valid for a perfect fluid and written with strain tensor invariants (and solid-like Landau moduli), to the development Eq. (3) of the isotropic solid.[10] An interpretation is equally given in terms of invariants: I_2 and I_3 could be seen as relative to shear deformation. This analogy also establishes relations between third- and fourth-order moduli. Due to the difference of state variable used in the equation of state of fluids (density variations) and used in the stress-strain relation of solids (strain tensor), first- and second-order coefficient A_l and B_l of the fluid are linked to third-order Landau moduli (of the solid-like development).

In fluids and solids, incompressibility can be commonly expressed by: $(\rho/\rho_0)^2 = 1 = \text{Det}(F_{ij})^2$, where $F_{ij} = \partial x_i / \partial a_j$ is the transformation tensor. The dilatation tensor C_{ij} ($C = \bar{F}^T \bar{F}$) respects a similar constraint: $III_C = \text{Det}(C_{ij}) = 1$, but cannot be used to develop the energy because it is not an infinitesimal quantity. This tensor is commonly used in hyperelasticity theory.[12]

Starting from the general form given by Eq. (3), the energy has been written in the simplified form by Hamilton et al.:[5,10]

$$W = W_0(\rho) + \mu(\rho)I_2 + \frac{A(\rho)}{3}I_3 + D(\rho)I_2^2. \tag{10}$$

To obtain this form, two equivalent methods can be used, both based on the following relations:

$$III_C = 1 + 2I_s - 4II_s + 8III_s \tag{11}$$

$$= 1 + 2I_1 - 2I_2 + 2I_1^2 + \tfrac{4}{3}I_1^3 - 4I_1I_2 + \tfrac{8}{3}I_3. \tag{12}$$

Since only I_2 and I_3 seem to be involved in shear energy, we can first express the invariant I_1 as combination of the other ones, assuming $III_C = 1$ in the previous relation:

$$I_1 = I_2 - I_1^2 - \tfrac{2}{3}I_1^3 + 2I_1I_2 - \tfrac{4}{3}I_3, \tag{13}$$

which becomes by recursion [substituting to I_1 its expression Eq. (13) until terms involving I_1 are insignificant]:

$$I_1 = I_2 - \tfrac{4}{3}I_3 + I_2^2. \tag{14}$$

Then, by substituting this expression of I_1, in the development Eq. (3), we obtain Eq. (10) with

$$D = \frac{\lambda}{2} + B + G. \tag{15}$$

With this method the reference value $W(\rho_0)$ is zero, and the three elastic moduli are those of the equilibrium state: $\mu(\rho_0)$, $A(\rho_0)$, and $D(\rho_0)$. In the course of calculation, supposing to know the order of magnitude of each term, the fourth-order terms involving moduli C, E, F, H, and higher order term combination of invariants have been neglected.

Always starting form Eq. (11), it is possible to express the dependence of this moduli with the relative volume variation III_c and I_2 and I_3. Solving Eq. (11) for I_1 gives a unique real solution depending on I_2 and I_3 at different orders and on III_c,

$$I_1 = \frac{1}{2}(-1 + \zeta_1 + \sqrt{2 + \zeta_2})^{1/3} - \frac{2(1/4 - I_2)}{(-1 + \zeta_1 + \sqrt{2 + \zeta_2})^{1/3}}$$
$$- \frac{1}{2}, \quad \text{with } \zeta_1 = 3III_C - 8I_3 \quad \text{and } \zeta_2 = -12I_2 + 48I_2^2$$
$$- 64I_2^3 - 6III_C + 16I_3 + 9III_C^2 - 48III_CI_3 + 64I_3^2. \tag{16}$$

This solution is developed using successive series expansions, considering the combination of invariants I_2 and I_3 as infinitesimal quantities and are consistent with a fourth-order approximation.

Then, like previously, the expression of I_1 is substituted into Eq. (3) or Eq. (4) to obtain a development whose coefficients are functions of elastic moduli and combination of I_2, I_3 and III_c. Finally, taking $III_C = 1$ and neglecting high order terms leads to Eq. (10). The coefficients obtained by the second way are not produced here, because of their complexity and their weak interest in practice. However, such calculation shows explicitly the dependence of coefficients μ, A, and D on the other elastic constants and invariants. This dependence characterizes the degree of incompressibility, and could give a relation between the approximation adopted on deformation and the resulting approximation conceded on shear moduli.

It could be simply seen through the dependence of D on λ that it is not strictly convenient to assign to A or D the physical meaning of shear moduli. In fact the incompressibility is rigorously taken into account by a constraint on the strain tensor ($III_C = 1$). Taking into account the incompressibility with a constraint on the elastic moduli is equivalent to finding a set of elastic moduli that cancel certain components of the stress tensor (those corresponding to transformations at nonconstant volume).

Polarization	Transverse	Longitudinal
Elastic moduli (Pa)	$\mu \approx 5000$	$\lambda = 2.5 \times 10^9$
Velocity (m/s)	2	1500
Viscosity (Pa s)	$\eta \approx 0.6$	0.005
Frequency (Hz)	100	4.3×10^6
Attenuation (Np/m)	13.4	0.025
Nonlinearity parameter	20	3.5
Wavelength (mm)	20	1.5

C. Modified Burgers equation

In this section we consider a plane transverse wave of wave number k propagating in the z direction (corresponding to the direction a_1 in Sec. II B) in an isotropic medium (the displacement, denoted u, is polarized along directions $a_2 \equiv x$ or $a_3 \equiv y$). In this simpler case, the viscous stress can be added and is assumed to be

$$\tau_{ik} = 2\eta \left[\frac{dS_{ik}}{dt} - \frac{1}{3}\partial_{ik}\frac{dS_{ll}}{dt} \right] + \zeta\partial_{ik}\frac{dS_{ll}}{dt}, \qquad (17)$$

where η and ς are the shear and volumic viscosity coefficients. This corresponds to a viscoelastic Voigt's model. Equation (17) and the energy development Eq. (10) simplify assuming a pure shear wave polarized along the z axis. The wave equation can then be rewritten in its nondimensional form ($\tilde{u} = u/u_0$, $\tilde{z} = kz$ and $\tilde{t} = \omega t$):

$$\frac{\partial^2 \tilde{u}}{\partial \tilde{z}^2} - \frac{\partial^2 \tilde{u}}{\partial \tilde{t}^2} = -\beta_T M^2 \frac{\partial^2 \tilde{u}}{\partial \tilde{z}^2}\left(\frac{\partial \tilde{u}}{\partial \tilde{z}}\right)^2 - \frac{\eta\omega}{\mu}\frac{\partial^3 \tilde{u}}{\partial \tilde{z}^2\,\partial \tilde{t}} \qquad (18)$$

with

$$\beta_T = 3\left(1 + \frac{\dfrac{A}{2} + D}{\mu}\right).$$

This equation is simplified by first introducing the (dimensionless) retarded time $\tilde{\tau} = \omega\tau = \tilde{t} - \tilde{z}$, and then substituting to $\tilde{z}$ the slow coordinate $M^2\tilde{z}$:

$$\frac{\partial^2 \tilde{u}}{\partial \tilde{\tau}\,\partial \tilde{z}} = \frac{\beta_T M^2}{2}\frac{\partial^2 \tilde{u}}{\partial \tilde{\tau}^2}\left(\frac{\partial \tilde{u}}{\partial \tilde{\tau}}\right)^2 + \frac{\eta\omega}{2\mu}\frac{\partial^3 \tilde{u}}{\partial \tau^3}. \qquad (19)$$

This cubic nonlinear equation is valid if $\beta_T M^2 \ll 1$, i.e., by neglecting fourth-order terms compared to both terms (cubic nonlinearity and viscosity) on the right-hand side of Eq. (19). The characteristic values, in our experimental conditions, are presented in Table II. The previous approximation is not always well verified, since $\beta_T M^2$ can reach a value equal to 0.5, but remains conceivable in most cases.

Finally, the so-called modified Burgers equation[13] is expressed with the particle velocity $v = \partial u/\partial \tau$ (and can be compared to the longitudinal case of quadratic nonlinearity[14]):

$$\frac{\partial v}{\partial z} = \frac{\beta_T}{2V_T^3}v^2\frac{\partial v}{\partial \tau} + \frac{\eta}{\rho_0 V_T^3}\frac{\partial^2 v}{\partial \tau^2}. \qquad (20)$$

Neglecting the dissipation term allows one to solve this equation with classical methods (implicit solution for simple
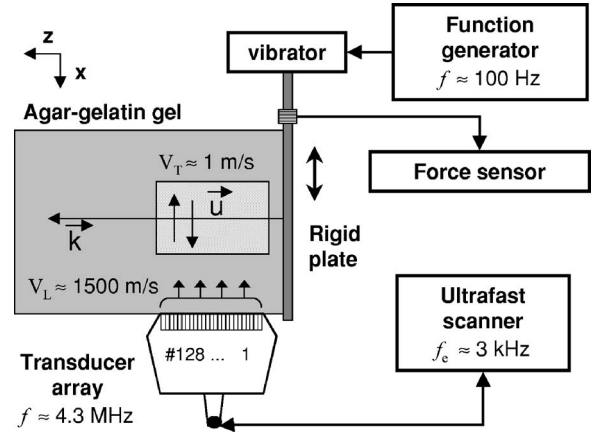
FIG. 1. Experimental setup: The vibrator and the rigid plate generate low frequency ($f=50–500$ Hz) shear waves propagating at $V_T \approx 2$ m/s. The ultrafast scanner records the backscattered longitudinal waves ($V_L = 1500$ m/s), received on the transducer array (central frequency $f_c = 4.3$ MHz). The correlation of these signals (acquired with the repetition rate of the scanner) gives the particle velocity in the (x,z) plane (spatial resolution of 0.3 mm along z and 0.5 mm along x), sampled at $f_e = 3$ kHz.

boundary conditions,[13–15] or successive approximations for weak nonlinearity). The shock distance is given by

$$z_c = \frac{2V_T^3}{\beta_T \omega_0 v_0^2}, \qquad (21)$$

and the attenuation coefficient by

$$\alpha = \frac{\eta\omega_0^2}{2\rho_0 V_T^3}. \qquad (22)$$

Nevertheless, as in gels the attenuation length $L_\alpha = 1/\alpha$ is significant, such a simplification is no longer relevant. A numerical model, transposed from the quadratic (longitudinal) case,[16] has thus been developed for this cubic nonlinear equation.

III. EXPERIMENTAL STUDY

A. Experimental setup, harmonic generation

The first experiments have been reported in a previous paper[4] dealing with harmonic generation of transverse shear waves. As shown in Fig. 1, a rigid plate (11×17 cm^2) is placed on the surface of an agar-gelatin phantom ($14 \times 14 \times 30$ cm^3). To assure a good mechanical contact, the plate is introduced during gel solidification. The phantom is composed of gelatin (5% in mass) and agar (3% in mass) mixed with hot water (between 50 and 80 °C) and left to rest one night at around 5 °C for gel solidification. Experiments are carried out at ambient temperature (between 20 and 25 °C).

The plate is fixed to a vibrator (Bruël&Kjaer model 4809). The excitation signal, composed of a few cycles (harmonic or multifrequency), is controlled by a function generator, amplified, and then applied to the vibrator. A force sensor records the source signal allowing one to control its linearity. The vibration of the plate ($f=100$ Hz) generates a shear wave of displacement vector **u** parallel to x axis propagating along z direction.
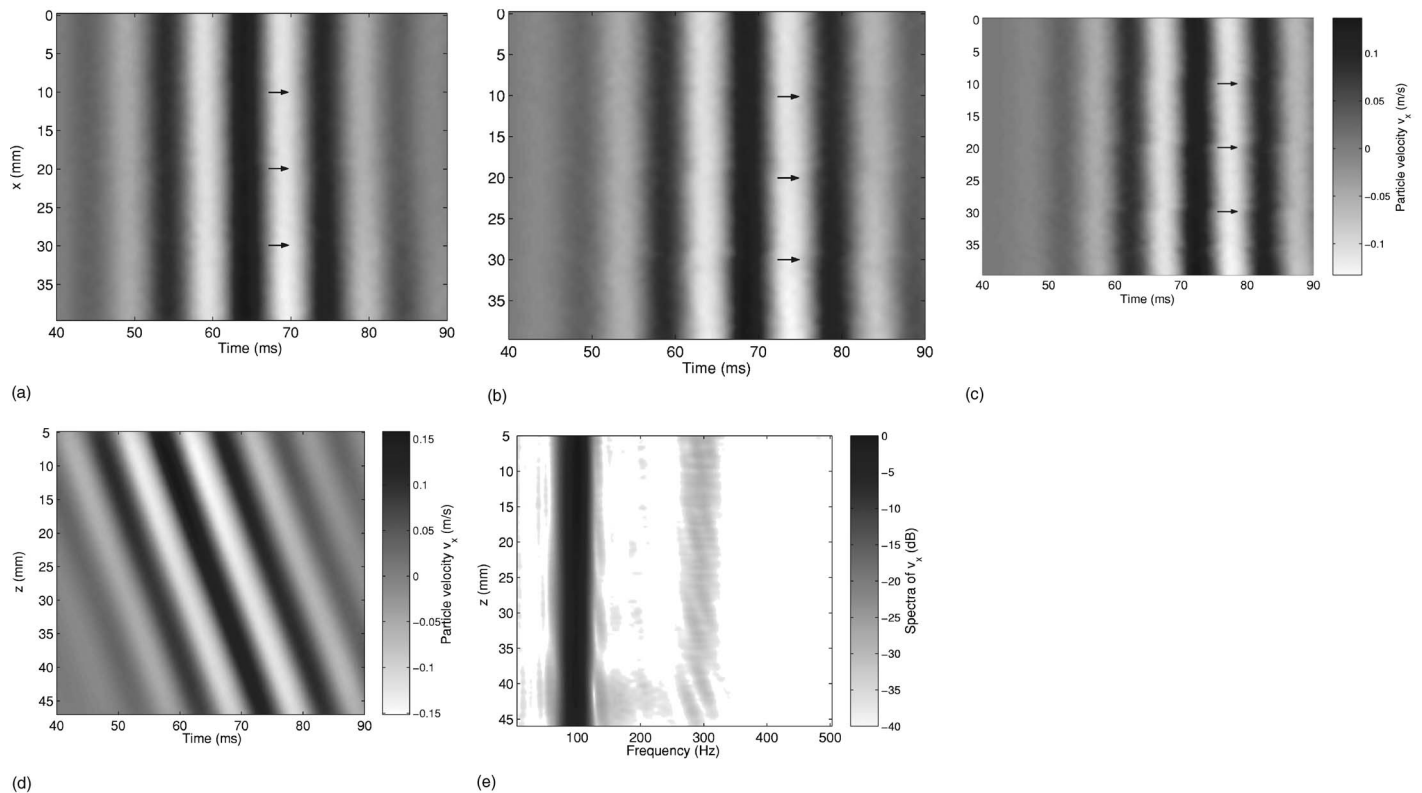
FIG. 2. Top (a), (b) and (c) Plane shear wave (v_x component) along x three propogation distances. Bottom Wave form of the shear wave $v_x(z)$ along the propogation direction (d).

The particle velocity is detected using an ultrafast scanner.[17] It consists in a standard ultrasound transducer array composed of 128 elements of central frequency 4.3 MHz, driven by a specific electronic device (128 independent channels, having their own DAC and memory, with a 50 MHz sampling frequency). In order to observe the velocity field $v_x(x,z,t)$, i.e., the temporal evolution of particle velocity in the (x,z) plane, the emitting surface of the transducer array is placed perpendicular to the polarization $\mathbf{u}(x)$ and with its larger dimension along the propagation axis z. The transverse dimension of the image (along z) is imposed by the length of the transducer array (4 cm). The depth along the x axis can be chosen between a few millimeters and a few centimeters. The zone is placed in the middle of the gel (relatively to x) to avoid boundary effects.

During the propagation of the low frequency shear wave coming from the plate, the ultrafast scanner emits plane longitudinal waves (without beamforming) at a repetition rate of 3 kHz (which is the sampling frequency of the low frequency wave). The beamforming is processed in receiving mode and finally the displacements are computed from cross-correlation algorithms.

In Fig. 2, the particle velocity fields $v_x(x,t)$ as function of time are represented on color scale images. The same wave form is observed in the transverse direction (x axis), which shows the planar character of the wave. A delay corresponding to time propagation can be observed between depth $z=10, 26.5$, and 43 mm in Fig. 2. On the bottom, Fig. 2(d) shows the particle velocity $v_x(z,t)$, averaged over x, from which we measure the speed of the shear wave (V_T

$=2.1$ m/s) and the attenuation coefficient ($\alpha=8.6$ Np/m), using, respectively, the phase and amplitude evolutions versus propagation distance. The viscosity coefficient is deduced from attenuation assuming a viscoelastic Voigt's model. Since nonlinear effects affect the amplitude of the fundamental component (100 Hz), the measurement of attenuation has to be done at a low excitation level.

At a higher excitation level ($v_0=0.5$ m/s and $M=0.24$), the sensor placed between the vibrator and the plate allows one to check that no harmonic is present at the source, as shown in Fig. 3(a) by the wave form and its spectrum. The wave form distortion, presented in Fig. 3(b), differs significantly from a nonlinearity distorted longitudinal wave. Here, only odd harmonics at 300 and 500 Hz are present in the spectrum in Fig. 3(b), which is a characteristic of cubic nonlinear effects affecting transverse waves. It may be pointed out that coupling effects, as described in Refs. 7, 9, and 18, are too small to give rise to a longitudinal component centered at the double frequency. The measured field $v_z(x,z,t)$, obtained by placing the transducer array in front of the plate, does not contain any 200 Hz component. This effect should be greater in "hard solids" because of the favorable ratio of velocities, but remains relative to the incident shear wave amplitude.

The nonlinear coefficient β_T, relative to shear waves, has been estimated, by fitting experimental and numerical results. The wave form and spectrum of the particle velocity, represented in Fig. 3(b), are fitted near the source. Then the parameter is corrected to fit the evolutions of the fundamental (100 Hz) and harmonics (300 and 500 Hz) amplitudes as

FIG. 3. Comparison between experiment and simulation in the case of harmonic generation. The signal shown in (a), obtained from the force sensor, is used as source signal for the calculation. The input parameter of the simulation are listed in Table III.

TABLE III. Input parameters of the numerical model of the modified Burgers equation corresponding to the harmonic generation experiment presented and to the nonlinear interaction case.

Experiment	Harmonic generation	Two-wave interaction
Shear modulus (Pa)	4400	6300
Speed (m/s)	2.1	2.4
Particle velocity (m/s)	0.5	0.62
Mach	0.24	0.26
Viscosity (Pa s)	0.45	0.7
Frequency (Hz)	100	100, 140
Attenuation (Np/m)	8.6	9.1, 17.8
Nonlinearity parameter	17	20
Wavelength (mm)	21	24, 17
Shock distance (mm)	7	5.7, 4.1

(2) The uncertainty linked to the source amplitude (which is estimated using the particle. velocity given by these first elements).

(3) The measurement of the attenuation coefficient that is not as accurate ($\pm 10\%$) as the linear velocity measurement ($\pm 1\%$). However, a relatively good agreement is found.

Similar comparisons have been performed using the more general case of two wave interaction at frequencies f_1 and f_2 (which are reduced to the harmonic generation when $f_1 = f_2$). Such an interaction, involving cubic nonlinearity, will give rise to numerous odd frequency components.

B. Plane wave interaction

The vibrator is now excited by a superposition of two sinusoidal components. Different frequency ratios have been tested. The more significant results have been obtained for two frequencies close to each other because of the strong attenuation of agar-gelatin phantoms. Choosing the frequencies of the excitation signal in the lower part of the bandwidth allows one to observe harmonic components at sum and difference frequencies. In fact, higher frequencies are too quickly distorted and attenuated and the use of lower frequencies is limited by the burst duration.

The signal applied to the vibrator is composed of a superposition of sinusoidal components at frequencies $f_1 = 100$ Hz and $f_2 = 140$ Hz. The signal measured by the force sensor and its spectrum are represented in Fig. 5(a). This wave form, as in the previous part, is used as an input in the simulation while its amplitude is estimated from the particle velocity measured in front of the first element of the array. On the spectrum, the fundamental frequencies at 100 and 140 Hz are clearly observed. The comparison between experiment and simulation is shown in Fig. 5(b).

A closer look to the spectra as function of depth represented on a gray scale image in Fig. 6 is necessary to compare the results and to highlight the created harmonics. The sum ($2f_1 + f_2$, $2f_2 + f_1$) and odd frequencies ($3f_1, 3f_2$) are clearly observed in experiments and in simulations. Unfortunately, difference frequencies ($2f_1 - f_2$, $2f_2 - f_1$), although present in the spectrum, have the same level as sidelobes of

a function of propagation distance (Fig. 4). The value of $\beta_T = 17$ (the parameters entered in the model are listed in Table III) is obtained with an uncertainty of the order of $\pm 10\%$. The main sources of error are:

(1) The approximative knowledge of the distance between the source (plate) and the first element of the array.

FIG. 4. Evolution of fundamental component and third and fifth harmonics vs propagation distance, experiment (dotted lines), and simulation (solid lines).

Jacob *et al.*: Nonlinear shear wave interaction

FIG. 5. Comparison between experiment and simulations. The signal shown in (a), obtained from the force sensor, is used as source signal for the calculation. The input parameters for the simulation are listed in Table III.

the emitted signal. These sidelobes are due to the finite duration of the emitted signal. The latter is imposed by the size of the sample: For a longer emission signal, the echo coming from the side opposite to the plate interferes with the direct traveling wave.

Another way of testing the model is to change the amplitude of one of the two primary components. At a given depth ($z=10$ mm), relative amplitudes of the most significant harmonics are plotted in Fig. 7 versus increasing amplitudes of the component centered at f_2. The component at $2f_2+f_1$ involving two times this frequency varies faster than the other one ($2f_1+f_2$). This follows qualitatively the results of a perturbation method. Such method, also called successive approximation and equivalent to series development, could be applied to the lossless wave equation. This consists in considering the nonlinear contribution as a correction of the linear solution (see Chap. 10 in Ref. 14), which gives the $2f_2+f_1$ and $2f_1+f_2$ component amplitudes respectively proportional to $k_1 k_2 U_1 U_2^2$ and $k_1 k_2 U_1^2 U_2$. U_1 and U_2 are the amplitudes, and k_1, k_2 the wave numbers, of the components f_1 and f_2 at the source. The quadratic dependence on U_2 is not checked in the experiment, which confirms that in gels the attenuation cannot be neglected and the level of nonlinearity is too high to use such solutions.

The numerical model, which takes into account attenuation, is in good agreement with this experiments. The input
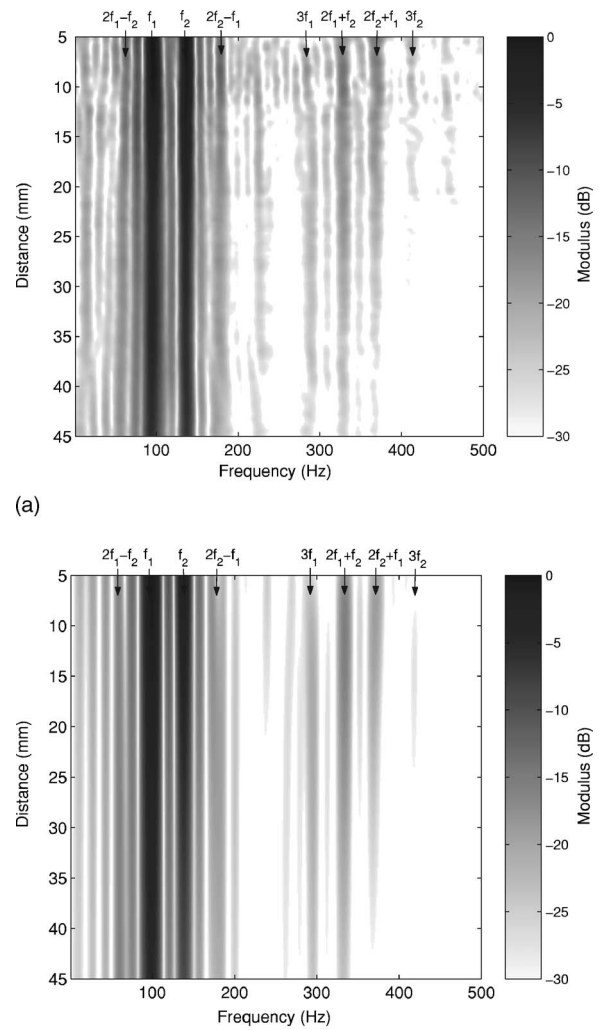
FIG. 6. Fourier spectra (modulus in decibels) of the v_x component vs propagation distance z: (a) experiment and (b) simulation.

parameters for the model are recalled in Table III. The adjustment of the experimental and theoretical data has been done using the wave form (and spectrum) and the evolution of the most significant harmonics versus distance. The use of wave interaction allows one to take full advantage of the large bandwidth of ultrafast imaging. An accurate measurement of the source amplitude would lead to a better agreement with all spectral components and an accurate estimation of the nonlinearity coefficient.

The possibility to have a nonlinear behavior for attenuation (amplitude dependent) is omitted in the theoretical approach (the quadratic dependence of the attenuation coefficient versus frequency follows from the hypothesis of a viscous stress tensor proportional to deformation rate). Such comparison between experiments and simulations could be used to test nonlinear viscoelastic models and improve the measurement accuracy.

C. Nonplane wave interaction

The purpose of most of the experiments presented in this paper is to quantify the nonlinearity of agar-gelatin phantoms, which need a simple model depending on a few pa-

rameters. This is the case for plane wave experiments exposed up to now. The understanding of the nonlinear propagation of nonplane shear waves is also of interest, since it corresponds to simpler experimental conditions.

The unidimensional model exposed in Sec. II C and the simplified energy development obtained in Sec. II B are not adapted to the description of diffraction effects in nonlinear regime. Before the statement of experimental results, we recall the solutions obtained by Zabalotskaya.[7] The Gaussian beam resolution adopted in Ref. 7 is not strictly transposable to the one-dimensional excitation conditions used in this part, but it illustrates the principal features of finite size source effects in the nonlinear regime.

Diffraction and/or coupling effects can be taken into account by starting from a general form of the stress tensor [Eq. (3)]. With initially L wave at the source, the vectorial displacement field $\mathbf{u}(a_1, a_2, a_3, t)$ is written as the sum of its component, adequately weighted by their order of magnitude:

$$\mathbf{u} = M u_3 + M\sqrt{M}(u_1 + u_2), \tag{23}$$

where u_1, u_2, and u_3 are functions of the retarded time $\tau = t - a_3/V_L$, and of the slow coordinates $a_3' = Ma_3$, $a_1' = \sqrt{M}a_1$ and $a_2' = \sqrt{M}a_2$ (the displacement is a slowly varying function of space coordinates).

The same method is applied to an initially transverse component. From the solution written as

$$u = \sqrt{M}(u_1 + u_2) + M u_3, \tag{24}$$

follows a nonlinear wave equation taking into account both diffraction and nonlinearity:

$$\frac{\partial^2 u_i}{\partial \tau \partial z} - \frac{V_T}{2}\Delta_\perp u_i - \frac{F}{2\rho_0 V_T^5}\frac{\partial}{\partial \tau}\left(\frac{\partial u_i}{\partial \tau}\frac{\partial u_j}{\partial \tau}\frac{\partial u_j}{\partial \tau}\right) - \frac{\eta}{2\rho_0 V_T^3}\frac{\partial^3 u_i}{\partial \tau^3}$$
$$= \frac{\mu + A/4}{2\rho_0 V_T^3}\left[\frac{\partial}{\partial \tau}\left(\frac{\partial u_j}{\partial a_i}\frac{\partial u_j}{\partial \tau} + \frac{\partial u_i}{\partial a_j}\frac{\partial u_j}{\partial \tau} - 2\frac{\partial u_i}{\partial \tau}\frac{\partial u_j}{\partial a_j}\right)\right.$$
$$\left. + \frac{\partial}{\partial a_j}\left(\frac{\partial u_i}{\partial \tau}\frac{\partial u_j}{\partial \tau}\right) - \frac{\partial}{\partial a_i}\left(\frac{\partial u_i}{\partial \tau}\frac{\partial u_j}{\partial \tau}\right)\right] \tag{25}$$

with i, $j = 1$, 2 and

$$F = \frac{\lambda}{2} + B + G + \frac{A}{2} + \mu = \frac{\mu\beta_T}{3}.$$

The left-hand terms (first line) correspond, respectively, to linear propagation and diffraction, cubic nonlinearity, and attenuation. This equation reduces to Eq. (20) considering plane waves: The second term becomes null as well as some part of the nonlinear term (coupling terms of third order). The right-hand side of the equation contains transverse derivatives of the T components. This second-order terms are nonzero because of diffraction, and weighted by a second-order modulus. Their physical meaning could be the nonlinear quadratic coupling of polarization due to the beam spreading (or finite size sources). The solution obtained by Zabalotskaya with Gaussian beam shows that in the cubic approximation the propagation of an initially pure shear wave gives rise to both longitudinal (noncumulative) and
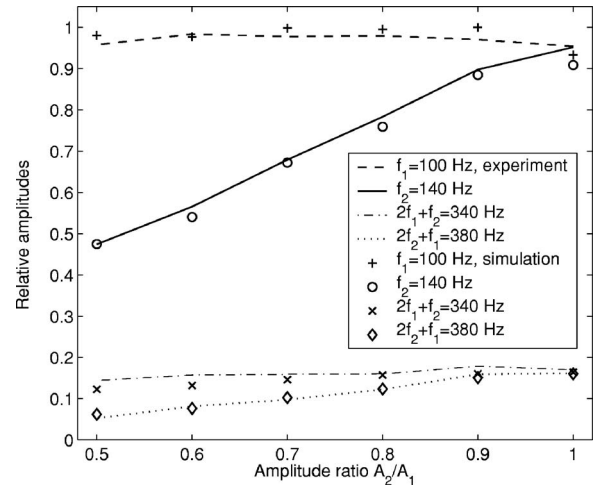
FIG. 7. Variation of the significant harmonics vs fundamental amplitude A_2 variations at a distance $z = 10$ mm from the source.

transverse components at double frequency. Thus, taking into account diffraction induces quadratic nonlinear effects on transverse waves.

The one-dimensional transient elastography technique[19] allows one to measure low frequency shear waves propagating in soft solids. It has already been applied to the measurement of the nonlinearity of gels using a static stress[3] (acoustoelasticity) and to the muscle anisotropy characterization.[1] As presented in Fig. 8, polarized shear waves are excited using a bar (3×80 mm^2) fixed to a vibrator. The shear waves are generated by pushing the bar on the gel surface, but the wave field emitted is composed of strains polarized both along and transversally to the direction of the vibrator motion. The monoelement transducer gives the component $v_z(z, t)$ of the particle velocity (longitudinal-like polarization), propagating at the speed of shear waves (2 m/s). The v_z component is the main component, but results from a complex deformation coupling the three components of the displacement in the near field of the source. As in the first part, the signal applied to the vibrator is composed of a superposition of sinusoidal components at frequencies $f_1 = 100$ Hz and $f_2 = 140$ Hz.

The harmonic generation experiment is presented in Fig. 9. No strictly longitudinal mode (i.e., wave propagating at 1500 m/s) is detected, because of the very different speeds
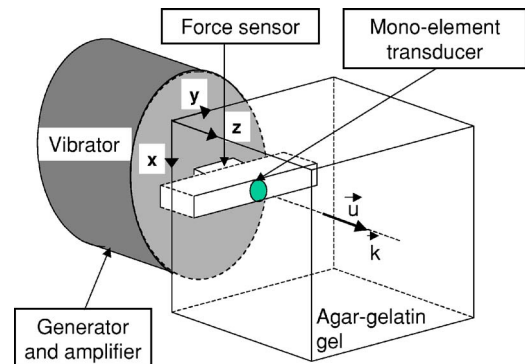
FIG. 8. (Color online) One-dimensional transient elastography (monoelement transducer) using a pushing bar as source.

Jacob *et al.*: Nonlinear shear wave interaction
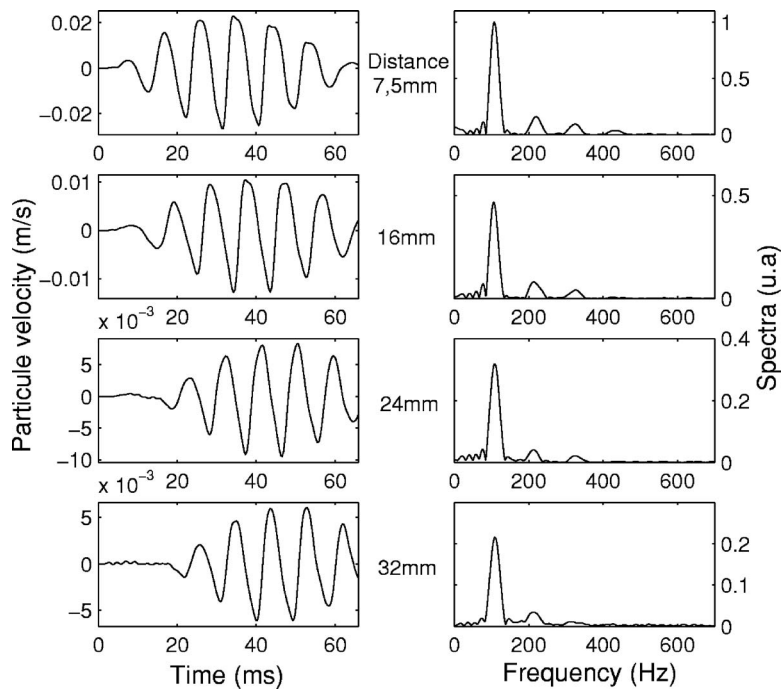
FIG. 9. Harmonic generation experiment with one-dimensional elastography showing the evolution, for successive propogation distances, of the wave form $v_z(z,t)$ (left column) and its spectra (right column).

of T and L waves [the compression induced by the bar motion propagates instantaneously (at 1500 m/s) relative to T waves (at 2 m/s)]. The wave form distortion is different from the one of pure shear waves presented in Sec. III A. This corresponds to the presence of a second harmonic in the spectrum.

The above-mentioned solution is not adapted to this type of source and cannot be compared to experimental results (one lateral dimension of the source is less than the shear wavelength), but gives an explanation for the nonlinear effect observed with such a source. In the linear case, the wave generated by the bar can be simulated using Green's functions,[20,21] i.e., using impulse responses taking into account the coupling in the near field. The second-order development of similar solutions, established by Jones and Kobett,[18] may be extended.

It is also possible, using two-dimensional elastography (128 elements transducer array) to characterize the v_x and v_y components. Because of the symmetry of the source the latter are null on the axis of the transducer. Apart from the axis z, v_x is the strongest component (≈ 0.2 m/s) whereas the v_y component is of the same order as v_z. Both projections of the displacement contain even as well as odd combination frequencies and have a penetration depth smaller than v_z.

Wave interaction has also been tested to confirm this observation. Figure 10 presents the spectral evolution of v_z along the propagation direction (z), obtained with a source excitation signal similar to the one plotted in Fig. 5(a). The most significant harmonics correspond to both quadratic ($2f_1, 2f_2, f_1 \pm f_2$) and cubic ($f_1 \pm 2f_2$) nonlinearity. With this representation one observes that the penetration depth is much smaller with the one-dimensional source: The nonlinear components reach their maximum at a 3 mm distance from the source and are quickly damped, because of viscosity and diffraction.

IV. CONCLUSION

The nonlinear elastic behavior of soft solid can be described using a simplified development of the elastic energy density, composed of one term at each order, weighted by the three elastic moduli μ, A, and D. The condition of quasi-incompressibility, imposed to write this development, is pertinent since we consider pure (plane) shear waves. The difference between compression and shear moduli justify this approximation and determine its order of magnitude.

Experiments of transient elastography give a synthetic view of these elastic properties: The ultrafast scanner exploits the fast (water-like) speed of L waves to measure the displacement of T waves, propagating at low velocity. Plane wave experiments illustrate well the cubic nonlinearity of
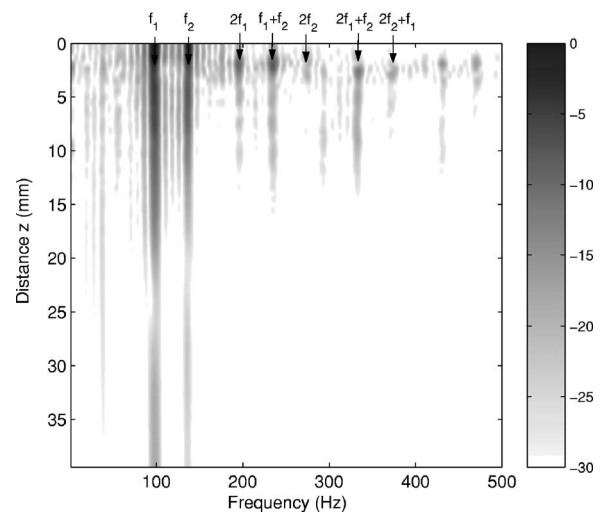
FIG. 10. Wave interaction with shear elasticity probe: evolution of the spectrum (decibels) of the v_x component vs propogation distances z. Both odd and even harmonics are present, those of significant amplitude are indicated above.

pure transverse waves. The numerical model of the modified Burgers equation is in good agreement with experimental results obtained for plane shear waves in both cases of harmonic generation and collinear wave interaction. The fitting of experimental and numerical result gives an estimation of the shear nonlinear coefficient of agar-gelatin phantoms. Surprisingly, the experiments confirm that, although expressed as function of compression moduli λ, B, G, the moduli D is of the order of magnitude of the so-called shear moduli μ and A. To our knowledge, no explanation has been given for this order of magnitude. As a practical consequence, shock distance is much more easily reached in soft solids than in "hard" solids with longitudinal waves (and several orders of magnitude between second- and third-order elastic moduli).

The shear elasticity probe is a simpler, and potentially applicable, experimental configuration. In this one-dimensional case, both harmonic generation and two wave interaction have also been tested. This experiment shows the role of diffraction in the nonlinear propagation of shear waves, the harmonic generation corresponding to quadratic nonlinearity. Results cannot be compared to a known model since it is necessary to take into account the coupling of T and L polarizations, i.e., the diffraction in the near field (a finite size of the source, like Gaussian profile, induces equally quadratic effect farther from the source). Further work has to be done to characterize the directivity in nonlinear regime.

ACKNOWLEDGMENTS

[1] J. L. Gennisson, S. Catheline, S. Chaffai, and M. Fink, "Transient elastography in anisotropic medium: Application to the measurement of slow and fast shear waves velocities in muscles," J. Acoust. Soc. Am. **114**, 536–541 (2003).

[2] J. D'hooge, B. Bijnens, J. Nuyts, J. -M. Gorce, D. Friboulet, J. Thoen, F. V. de Werf, and P. Suetens, "Nonlinear propogation effects on broadband attenuation measurements and its implications for ultrasonic tissue characterization," J. Acoust. Soc. Am. **106**, 1126–1133 (1999).

[3] S. Catheline, J. L. Gennisson, and M. Fink, "Measurement of elastic nonlinearity of soft solid with transient elastography," J. Acoust. Soc. Am. **114**, 3087–3091 (2003).

[4] S. Catheline, J. L. Gennisson, M. Tanter, and M. Fink, "Observation of shock transverse waves in elastic media," Phys. Rev. Lett. **91**, 43011–43014 (2003).

[5] M. F. Hamilton, Y. A. Ilinskii, and E. A. Zabolotskaya, "New formulation of the elastic energy density for soft tissue," J. Acoust. Soc. Am. **114**, 2436 (2003).

[6] L. Landau and E., Lifschitz, *Theory of Elasticity* (Butterworth-Heineman, London, 2002).

[7] E. A. Zabolotskaya, "Sound beam in a nonlinear isotropic solid," Sov. Phys. Acoust. **32**, 296–299 (1986).

[8] Z. A. Gol'dberg, "Interaction of plane longitudinal and transverse elastic waves," Sov. Phys. Acoust. **6**, 306–310 (1971).

[9] L. Zarembo and V. Krasil'nikov, "A nonlinear phenomena in the propagation of elastic waves in solids," Sov. Phys. Usp. **13**, 778-797 (1971).

[10] M. F. Hamilton, Y. A. Ilinskii, and E. A. Zabolotskaya, "Separation of compressibility and shear deformation in the elastic energy density," J. Acoust. Soc. Am. **116**, 41–44 (2004).

[11] X. Jacob, J. L. Gennisson, S. Catheline, M. Tanter, C. Barriere, D. Royer, and M. Fink, "Study of elastic nonlinearity of soft solids with transient elastography," Proc.-IEEE Ultrason. Symp. 660–663 (2003).

[12] *Continuum Mechanics for Engineers*, 2nd ed., edited by G. T. Mase and G. E. Mase (CRC Press, Boca Raton, London, 1999).

[13] E. A. Zabolotskaya, M. F. Hamilton, Y A. Ilinskii, and G. D. Meegan, "Modeling of nonlinear shear waves in soft solids," J. Acoust. Soc. Am. **116**, 2807–2813 (2004).

[14] *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, San Diego, 1998).

[15] R. W. Lardner, "Nonlinear effects on transverse shear waves in an elastic medium," J. Elast. **15**, 53–57 (1985).

[16] B. McDonald and J. Ambrosiano, "High-order upwind flux correction methods for hyperbolic conservation laws," J. Comput. Phys. **56**, 449–460 (1984).

[17] L. Sandrin, S. Catheline, M. Tanter, and M. Fink, "Shear modulus imaging using 2d transient elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 426–435 (2002).

[18] G. L Jones and D. R Kobett, "Interaction of elastic waves in an isotropic solid," J. Acoust. Soc. Am. **35**, 5-10 (1962).

[19] L. Sandrin, M. Tanter, J. L. Gennisson, S. Catheline, and M. Fink, "Shear elasticity probe for soft tissues with 1d transient elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 436–446 (2002).

[20] J. L. Gennisson, "The shear elasticity probe: A new tool for soft tissues investigation," Ph.D. thesis, University of Paris VI, Paris, France, 2003.

[21] D. C. Gakenheimer and J. Miklowitz, "Transient excitation of an half space by a point load traveling on the surface," J. Appl. Mech. **36**, 505–514 (1969).

Motion of a solid sphere in a viscoelastic medium in response to applied acoustic radiation force: Theoretical analysis and experimental verification

Salavat R. Aglyamov
Department of Biomedical Engineering, University of Texas at Austin, Austin, Texas 78712-0238
and The Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino,
Moscow Region, Russia 142290

Andrei B. Karpiouk
Department of Biomedical Engineering, University of Texas at Austin, Austin, Texas 78712-0238

Yurii A. Ilinskii and Evgenia A. Zabolotskaya
Applied Research Laboratories, University of Texas at Austin, Austin, Texas 78713-8029

Stanislav Y. Emelianov
Department of Biomedical Engineering, University of Texas at Austin, Austin, Texas 78712-0238

The motion of a rigid sphere in a viscoelastic medium in response to an acoustic radiation force of short duration was investigated. Theoretical and numerical studies were carried out first. To verify the developed model, experiments were performed using rigid spheres of various diameters and densities embedded into tissue-like, gel-based phantoms of varying mechanical properties. A 1.5 MHz, single-element, focused transducer was used to apply the desired radiation force. Another single-element, focused transducer operating at 25 MHz was used to track the displacements of the sphere. The results of this study demonstrate good agreement between theoretical predictions and experimental measurements. The developed theoretical model accurately describes the displacement of the solid spheres in a viscoelastic medium in response to the acoustic radiation force.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2774754]

I. INTRODUCTION

The mechanical properties and physiological state of biological tissues are often linked; as a result, pathological changes in tissues are usually related to changes in their biomechanical parameters.[1–4] Further, a noninvasive assessment of the mechanical properties of tissue is often needed in surgical and therapeutic practices. Recent studies have shown that an acoustical radiation force can be used in many biomedical applications including ophthalmology,[5–7] detecting and characterizing lesions,[8] the imaging of arteries with calcification,[9] monitoring thermal lesions during high-intensity focus ultrasonic treatment,[10] screening muscle condition,[11] the detecting foreign objects in the body,[12] etc.

An acoustic radiation force arising from absorption of an ultrasound wave can be used in remote palpation to either directly assess local mechanical properties of tissue,[13–18] or to produce excitation[19–22] in shear wave elasticity imaging. If an acoustic inhomogeneity is present inside the tissue, the radiation force can be generated through reflection of the ultrasound wave. For example, a harmonic, low-frequency radiation force was applied to solid spheres of various sizes.[23] The spheres were embedded in tissue-like phantoms to estimate the resonance frequencies of spheres and their dependence on the viscoelastic properties of the media. A laser vibrometer was used to measure the vibration speed of the sphere. It was demonstrated that the method can be used to determine the local material properties of the medium surrounding the sphere. Also, the acoustic radiation force was applied to a gas bubble produced by a femtosecond laser in tissue-like phantoms.[6,7] It was shown that the displacement of the bubble is affected by the elastic modulus of the surrounding gel.

In our previous studies we developed a general approach to estimate the displacement of rounded objects (specifically, gas bubbles and solid spheres) in a viscoelastic medium in response to applied acoustic radiation force.[24] The developed model could potentially help assess mechanical properties of tissues using microbubbles as acoustic targets. The first step of the validation process, and the goal of the current study, is to theoretically and experimentally investigate the dynamic behavior of a rigid sphere in a viscoelastic medium under a radiation force of short duration (less than 10 ms). A model to describe the motion of a solid sphere was derived and solved both analytically and numerically. Further, experimental studies using tissue-like gelatin-based phantoms with embedded solid spheres were conducted to verify the developed model. The experiments were performed with spheres of different sizes and densities. The motion of the sphere was induced by acoustic pulses of different durations. The elastic properties of the phantoms were varied. In all cases, there was a good agreement between theoretical predictions and experimental measurements. The results of our study strongly suggest that the developed theoretical model pre-

dicts the dynamic behavior of a solid sphere in a viscoelastic medium. Consequently, this theoretical model can be used to assess the mechanical properties of the medium surrounding the sphere.

II. THEORY

In our previous studies we considered the displacements of a solid sphere and a gas bubble in a linear viscoelastic medium, under an external, time-dependent, acoustic radiation force.[24] Theoretical results were obtained for compressible and incompressible media. However, as most soft tissues are nearly incompressible,[25] we consider an incompressible medium only. The equation of motion for an incompressible medium is

$$-\nabla P + \mu \nabla^2 \mathbf{U} + \eta \nabla^2 \frac{\partial \mathbf{U}}{\partial t} = \rho \frac{\partial^2 \mathbf{U}}{\partial t^2} \qquad (1)$$

where $\mathbf{U}$ is a displacement vector, μ and η are shear elastic and viscous coefficients, P is internal pressure, ρ is medium density, and t is time.

We assume that the sphere is placed in an axisymmetric field of the radiation force. The polar axis of the spherical system of coordinates is along the force vector (i.e., an angle θ is between a radius vector and displacement), $\mathbf{U} = (U_r, U_\theta, 0)$ and the boundary conditions at the surface of the rigid solid sphere of radius R are

$$U_r(R, \theta) = U \cos \theta, \quad U_\theta(R, \theta) = -U \sin \theta, \qquad (2)$$

where U is the displacement of the sphere. The external force applied to the displaced spherical object is[26]

$$F^{(\text{ext})} = -2\pi R^2 \int_0^\pi (\sigma_{rr} \cos \theta - \sigma_{r\theta} \sin \theta) \sin \theta \, d\theta, \qquad (3)$$

where σ_{rr} and $\sigma_{r\theta}$ are stress tensor components at the surface of the object:

$$\sigma_{r\theta} = \left(\mu + \eta \frac{\partial}{\partial t}\right) \left(\frac{\partial U_\theta}{\partial r} - \frac{U_\theta}{r} + \frac{1}{r} \frac{\partial U_r}{\partial \theta}\right),$$

$$\sigma_{rr} = -P + 2\left(\mu + \eta \frac{\partial}{\partial t}\right) \frac{\partial U_r}{\partial r}. \qquad (4)$$

Because the radial component of the strain vanishes ($\partial U_r / \partial r = 0$) on the surface of the sphere ($r = R$), Eq. (3) can be rewritten using only the shear component of stress tensor and the internal pressure:

$$F^{(\text{ext})} = 2\pi R^2 \int_0^\pi (P \cos \theta + \sigma_{r\theta} \sin \theta) \sin \theta \, d\theta. \qquad (5)$$

As shown previously,[24] the displacement of the sphere U_0 in a static case is

$$U_0 = \frac{F^{(\text{ext})}}{6\pi\mu R}, \qquad (6)$$

where $F^{(\text{ext})}$ is an external force applied on the surface of a rigid sphere (e.g., acoustic radiation force). This equation also describes the displacement of the sphere in a viscoelastic medium if constant external force is applied over a long
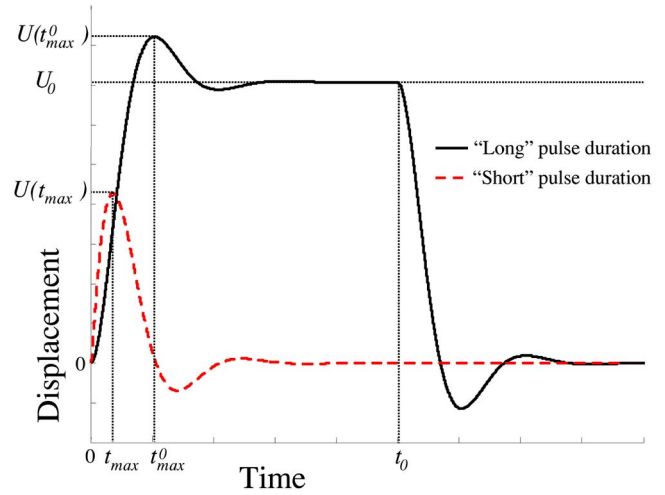
FIG. 1. (Color online) Schematic representation of the displacements of a dense sphere ($\beta > 5/8$) embedded in a purely elastic medium in response to "long" and "short" pulse durations (not to scale). If the duration of the acoustic radiation force pulse t_0 is sufficiently long, the displacement reaches its maximum value $U(t_{max}^0)$ at $t_{max}^0 < t_0$ and then the sphere oscillates around its equilibrium position U_0 until the pulse is ended and the sphere returns and oscillates around the original position.

period of time. Clearly, the sphere displaced by U_0 is in equilibrium position when external force is compensated for by the elastic forces from the medium.

The equation for the sphere motion (1) in the frequency domain is given by

$$-\nabla p + (\mu - i\omega\eta)\nabla^2 \mathbf{u} + \rho\omega^2 \mathbf{u} = 0, \qquad (7)$$

where p and $\mathbf{u}$ are the Fourier transforms of P and $\mathbf{U}$, ω is an angular frequency, and force dependence on time is assumed to be proportional to $e^{-i\omega t}$. Taking into account the inertial forces, the equation that couples a spectral component of solid sphere displacement u_ω and the Fourier transform of external force $F_\omega^{(\text{ext})}$ is obtained:[24]

$$F_\omega^{(\text{ext})} = -M\omega^2 u_\omega + 6\pi\mu R u_\omega\left(1 - ikR - \frac{1}{9}k^2 R^2\right), \qquad (8)$$

where M is a mass of solid sphere and k is a complex wave number $k^2 = \rho\omega^2/(\mu - i\omega\eta)$.

For an infinitely light sphere, where $M = 0$, Eq. (8) becomes a well-known solution for the impedance of the sphere.[27] After applying inverse Fourier transform we obtain the sphere displacement U in time domain:

$$U(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} u_\omega e^{-i\omega t} d\omega. \qquad (9)$$

In the elastic case, where $\eta = 0$, Eq. (8) for solid sphere motion can be written in time domain:[24]

$$U + \frac{R}{c_t}\dot{U} + \frac{1}{9}(1 + 2\beta)\frac{R^2}{c_t^2}\ddot{U} = \frac{F^{(\text{ext})}}{6\pi\mu R}, \qquad (10)$$

where $c_t = \sqrt{(\mu/\rho)}$ is the shear wave speed and parameter $\beta = \rho_s/\rho$ is the solid sphere density ρ_s normalized by medium density ρ. Initial conditions for the differential equation (10) are zero.

Here, we consider that acoustic radiation force applied to a solid sphere is impulsive, i.e.,

$$F^{(\text{ext})} = \begin{cases} F_0, & 0 \leq t \leq t_0 \\ 0, & t > t_0, \end{cases} \tag{11}$$

where t_0 is the duration of the acoustic radiation force pulse and F_0 is amplitude of the force. The Fourier transform of this function is

$$F_\omega^{\text{ext}} = -\frac{iF_0}{\omega}(e^{i\omega t_0} - 1). \tag{12}$$

In the elastic case, after substituting Eq. (11) into Eq. (10), an analytical solution of Eq. (10) can be obtained:

$$U(t) = \begin{cases} \dfrac{F_0}{6\pi\mu R}\left(1 + \dfrac{\lambda_2}{\lambda_1 - \lambda_2}e^{(c_t/R)\lambda_1 t} - \dfrac{\lambda_1}{\lambda_1 - \lambda_2}e^{(c_t/R)\lambda_2 t}\right), & 0 \leq t \leq t_0 \\ \dfrac{F_0}{6\pi\mu R}\left((1 - e^{-(c_t/R)\lambda_1 t_0})\dfrac{\lambda_2}{\lambda_1 - \lambda_2}e^{(c_t/R)\lambda_1 t} - (1 - e^{-(c_t/R)\lambda_2 t_0})\dfrac{\lambda_1}{\lambda_1 - \lambda_2}e^{(c_t/R)\lambda_2 t}\right), & t \geq t_0 \end{cases} \tag{13}$$

where

$$\lambda_{1,2} = \frac{3}{2}\frac{-3 \pm \sqrt{5 - 8\beta}}{1 + 2\beta}$$

are the roots of an algebraic equation: $(1+2\beta)\lambda^2 + 9\lambda + 9 = 0$.

Solution (13) is a sum of two time-dependent, exponential functions where the real parts of the exponents are always negative. If normalized density β is smaller than $5/8$, the exponents are real and, therefore, the sphere first displaces from, and then returns to, the origin exponentially. However, if the sphere is sufficiently dense ($\beta > 5/8$), the exponents are complex and the sphere starts to oscillate with a linear frequency

$$\frac{3c_t}{4\pi R}\frac{\sqrt{8\beta - 5}}{1 + 2\beta}$$

defined by the imaginary part of the exponent (Fig. 1).

Initially, once the acoustic force is applied, the sphere will be gradually displaced from the origin (Fig. 1). If the duration of the acoustic radiation force pulse t_0 is sufficiently long, the displacement of the dense sphere ($\beta > 5/8$) reaches its maximum value $U(t_{\text{max}}^0)$ at $t_{\text{max}}^0 < t_0$, where

$$t_{\text{max}}^0 = \frac{2\pi R}{3c_t}\frac{1 + 2\beta}{\sqrt{8\beta - 5}}, \tag{14}$$

and

$$U(t_{\text{max}}^0) = \frac{F_0}{6\pi\mu R}(1 + e^{-3\pi/\sqrt{8\beta - 5}}). \tag{15}$$

Once the sphere reaches the maximum of the displacement and while the acoustic radiation force is still acting on the sphere, i.e., during $t_{\text{max}}^0 < t < t_0$ (Fig. 1), the dense sphere will oscillate around its equilibrium position U_0 defined by Eq. (6). For a light sphere ($\beta < 5/8$), no oscillations will be present. For infinitely long duration t_0 of the acoustic radiation force, as expected, Eq. (13) converges to Eq. (6) regardless of the normalized density of the sphere. Here the acoustic force is compensated for by the elastic response of the medium, and the displaced sphere is stationary as long as the external force is applied on the sphere.

To better understand the relationship between the displacement of the sphere and the properties of the sphere and the surrounding medium, consider a limiting case when the duration of the acoustic pulse t_0 is infinitely short. Using the Taylor expansion of the time-dependent exponential functions around zero, the solution of Eq. (13) for impulsive excitation can be obtained:

$$U(t) = \frac{F_0 t_0}{6\pi R^2\sqrt{\mu\rho}}\frac{\lambda_1\lambda_2}{\lambda_1 - \lambda_2}(e^{(c_t/R)\lambda_1 t} - e^{(c_t/R)\lambda_2 t}). \tag{16}$$

The displacement reaches its peak at the time t_{max}

$$t_{\text{max}} = \frac{R}{c_t}\frac{\log(\lambda_2/\lambda_1)}{(\lambda_1 - \lambda_2)}, \tag{17}$$

and the maximum displacement $U(t_{\text{max}})$ is

$$U(t_{\text{max}}) = \frac{F_0 t_0}{6\pi R^2\sqrt{\mu\rho}}\frac{\lambda_1\lambda_2}{\lambda_1 - \lambda_2}\left(\left(\frac{\lambda_2}{\lambda_1}\right)^{\lambda_1/(\lambda_1 - \lambda_2)} - \left(\frac{\lambda_2}{\lambda_1}\right)^{\lambda_2/(\lambda_1 - \lambda_2)}\right). \tag{18}$$

Note that $U(t)$ and t_{max} are real even if λ_1 and λ_2 are complex.

The previous analysis demonstrates a difference in maximum displacement between short and long durations of acoustic radiation force pulse. For example, Eq. (18) shows that for short impulses the maximum displacement is inversely proportional to the square root of shear elasticity and the square of the sphere radius. However, for long pulses, the displacement is inversely proportional to both shear elasticity and the radius, as is evident from Eq. (15). When the excitation time t_0 cannot be considered infinitely short, but $t_0 < t_{\text{max}}^0$, both the maximum of the displacement $U(t_{\text{max}})$ and the time to reach the maximum of the displacement t_{max} are functions of t_0. Generally, therefore, the relationship between the spatiotemporal behavior of the sphere and the properties of the sphere and the surrounding medium is more complex as described by Eq. (13). Equations (13)–(18) are useful in the quantitative analysis of the motion of the sphere in the

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Aglyamov *et al.*: Solid sphere motion in viscoelastic medium   1929

TABLE I. Parameters of the solid spheres used in the radiation force experiments.

	A	B	C	D	E	F	G
Material	Sapphire	Sapphire	Sapphire	Glass BK7	Glass LaSFN9	Plastic	Metal
Density (kg/m^3)	3980	3980	3980	2510	4440	700	9060
Diameter (mm)	0.5	1.0	1.5	3.0	3.0	3.0	3.0

elastic medium. However, in the experimental verification of the developed theoretical model, the viscosity should be taken into account.

In case of the viscoelastic medium, the displacement of the sphere can be written in an integral form by combining Eqs. (8), (9), and (12):

$$U(t) = -\frac{iF_0}{12\pi^2 R}\int_{-\infty}^{\infty}$$

$$\times \frac{(e^{i\omega t_0}-1)e^{-i\omega t}}{\omega(\mu - i\omega\eta)(1 - ikR - k^2R^2(1+2\beta)/9)}d\omega. \quad (19)$$

Viscosity reduces both the oscillations in sphere displacement and the displacement maximum. Integral (19) can be evaluated numerically. In fact, this equation was used to obtain all theoretical estimations presented in this article.

III. EXPERIMENTAL METHODS

To verify our theoretical model, the experiments were performed using rigid spheres of various diameters and densities embedded into gel-based phantoms of varying mechanical properties. The parameters of the seven spheres used in our study are listed in Table I. Three spheres (A, B, and C) had the same density but different diameters. The other four spheres (D, E, F, and G) had the same diameter but different densities.

An experimental setup is schematically shown in Fig. 2. A plastic tank with a pushing (1.5 MHz) transducer attached at the bottom was filled with 3% by weight gelatin (300
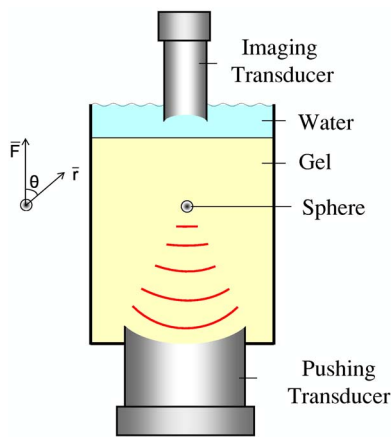
FIG. 2. (Color online) A schematic view of an experimental setup. A tank was filled with 3% by weight gelatin. A rigid sphere was embedded in the gel at the focal zone of a 1.5 MHz, single-element focused transducer attached to the bottom of the tank. To generate acoustic radiation force, short pulses were used. To monitor sphere position, a 25 MHz single-element transducer was placed at the top of the tank.

Bloom, type-A, Sigma-Aldrich, Inc., St. Louis, MO) to insure good contact between the phantom surface and the transducer. The resulting cubic phantoms with embedded rigid spheres were about 55 mm in linear size. The remaining portion of the tank was filled with water to provide contact between the imaging (25 MHz) transducer placed on the top of the tank and the phantom.

During phantom preparation, the same gelatin solution was used to prepare a cylindrical test sample of 35 mm diameter by 17 mm height. This sample was used to independently measure elastic properties of the phantom material. Both phantom and sample were constructed at the same time and underwent the same procedures (storage, experimental protocol, etc.) to minimize any possible differences in elastic properties. To estimate the shear elastic modulus of the samples, the uniaxial load–displacement measurements were performed using an In-Spec 2200 portable system (Instron, Inc., Norwood, MA). The control samples placed in a water tank were deformed axially up to 10% strain with 0.2 mm/s deformation rate. For each given temperature, the experiments were repeated at least five times to evaluate the accuracy of the measurements. In all measurements, the standard deviation did not exceed 15%.

Once a sphere was embedded into a phantom, the experiments were performed at the temperatures ranging from 4 to 22 °C. For gelatin, these temperature changes resulted in change of shear elasticity from 4 to 0.5 kPa (i.e., temperature and shear elasticity of gelatin are inversely proportional). Therefore, by varying the temperature, it was possible to change the elasticity of the surrounding material without changing the position of the sphere relative to the pushing transducer or the imaging transducer. The temperature of the phantom was controlled by a digital thermometer with a sensor embedded into the phantom about 10 mm away from the sphere.

A 1.5 MHz single-element focused transducer (Valpey Fisher Corp., Hopkinton, MA) was attached to the bottom of the tank to generate acoustic radiation force. The transducer has a diameter of 38 mm and a focal length of 44 mm. The function generator (Agilent 33250A, Agilent Technologies, Inc, Santa Clara, CA) was used to generate the sinusoidal signal applied through a radio frequency (rf) power amplifier (ENI model 240L, ENI, Rochester, NY) to the pushing transducer.

The position of the sphere was monitored using a 25 MHz, 6.35 mm diameter, single-element transducer with a focal length of 13 mm. The transducer was connected to a pulser/receiver operating at a 20 kHz pulse repetition rate (Panametrics 5910PR, Panametrics, Waltham, MA). The

Aglyamov *et al.*: Solid sphere motion in viscoelastic medium

pulse-echo signal was captured using an 8 bit, 500 MHz A/D data acquisition card (CompuScope 8500, GaGe Inc., Montreal, Canada).

Prior to the experiment, the foci of both the pushing and the imaging transducers were aligned at the location of the embedded sphere. The position of the sphere was tracked every 50 μs for up to 20 ms. The duration of acoustic radiation force pulse was varied from 0.33 to 10 ms. The displacement of the sphere was measured using a phase-sensitive, cross-correlation motion tracking algorithm.[28] The pushing pulse was applied 0.5 ms after the imaging sequence, consisting of repetitive pulse-echo firing, was initiated. The initial pulse-echo record captured prior to the radiation force pulse was used as a reference point for the cross-correlation procedure.

IV. RESULTS AND DISCUSSION

To verify the developed model describing the motion of a solid sphere in response to applied radiation force, a set of experiments was performed using spheres of different sizes and densities, various durations of acoustic radiation force pulses, and elastic properties of the surrounding material. In the theoretical analysis, Eq. (19) was used to compare the experimental and theoretical results. The amplitude of radiation force F_0 was obtained by fitting the theoretical predictions to experimental data. It was assumed that the shear viscosity of gelatin is 0.1 Pa s and does not depend on the temperature. This value provided the best agreement with experimental data and is consistent with literature data.[19,21]

The motion of the 0.5, 1.0, and 1.5 mm diameter spheres (spheres A, B, and C, correspondingly) in response to acoustic radiation force is presented in Fig. 3. The dotted vertical lines in Fig. 3 mark the beginning ($t=0.5$ ms) and the end of the 0.67 ms pushing pulse. The theoretical results in Fig. 3(a) were obtained assuming that radiation pressure in the focal zone of the excitation transducer is constant and equals approximately 732 Pa. This assumption is reasonable as the size of the sphere is smaller than or comparable to the 1.2 mm wide ultrasound beam in the focal zone of the 1.5 MHz transducer. As the radiation pressure is defined as the force divided by the surface area of the sphere, the pressure of 732 Pa corresponds to 0.57, 2.3, and 5.2 mN forces for the spheres A, B, and C, respectively.

Although the radiation force increases with the radius of the sphere, the elastic response of the material increases as well. The displacement of the sphere is larger for the larger sizes of the sphere. However, because in this experiment the duration of the acoustic pulse t_0 was not very short and $t_0 < t_{max}^0$, the relationship between displacement and the radius of the sphere is not defined by either Eq. (15) or Eq. (18) describing impulsive or long acoustic pulses, correspondingly. At the end of the pushing pulse the sphere continues to move until it reaches a maximum of the displacement. The independently measured shear elasticity of gelatin for all three phantoms was about 1300 Pa. This value was used in numerical analysis presented in Fig. 3(a). Figure 3(b) shows the results of the experimental studies for spheres embedded in three different phantoms with approximately the same
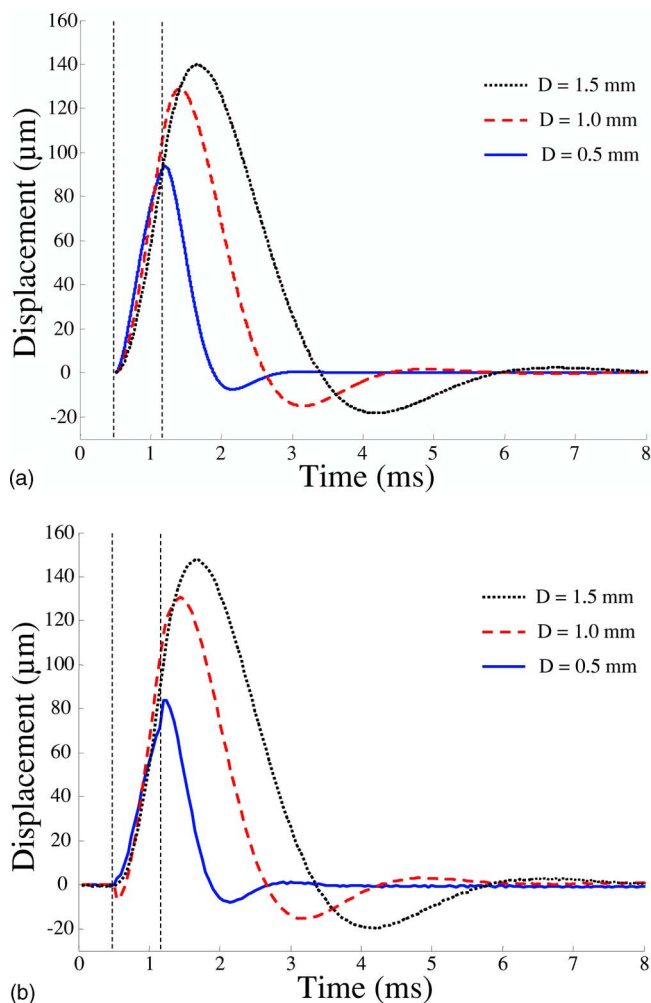
FIG. 3. (Color online) Displacements of 0.5, 1.0, and 1.5 mm diameter spheres (spheres A, B, and C in Table I) embedded in a viscoelastic medium. Acoustic radiation force excitation time was 0.67 ms, shear elasticity of media was measured and assumed to be 1300 Pa. (a) Theory: shear viscosity of media was 0.1 Pa s and the pressure on sphere surface was 732 Pa corresponding to 0.57, 2.3, and 5.2 mN forces acting on the spheres A, B, and C, correspondingly. (b) Experiment: note excellent agreement with theoretical model.

acoustical and mechanical properties. As expected, the displacements are larger for larger spheres. Clearly, the agreement between experimental results and theoretical predictions is excellent.

The motion of the sphere in response to radiation force pulses of different durations is presented in Fig. 4. The vertical dotted line in Fig. 4 indicates the beginning of the pulsed radiation force. In the experiments, a 1 mm diameter sphere (sphere B in Table I) was used. As the duration of the acoustic radiation force pulse increases, the displacement of the sphere initially increases and then it reaches the point of equilibrium where acoustic force is compensated for by the elastic response of the medium ($t_{max}^0 \approx 1.75$ ms). At the end of the acoustic pulse, this dense sphere ($\beta=3.98>5/8$) oscillates around the point of equilibrium. Similar oscillatory motion is present when the pushing pulse has ended and the sphere is returning to the original position. In addition, the time needed to reach the maximum of the displacement increases with the pulse duration. Again, the agreement be-
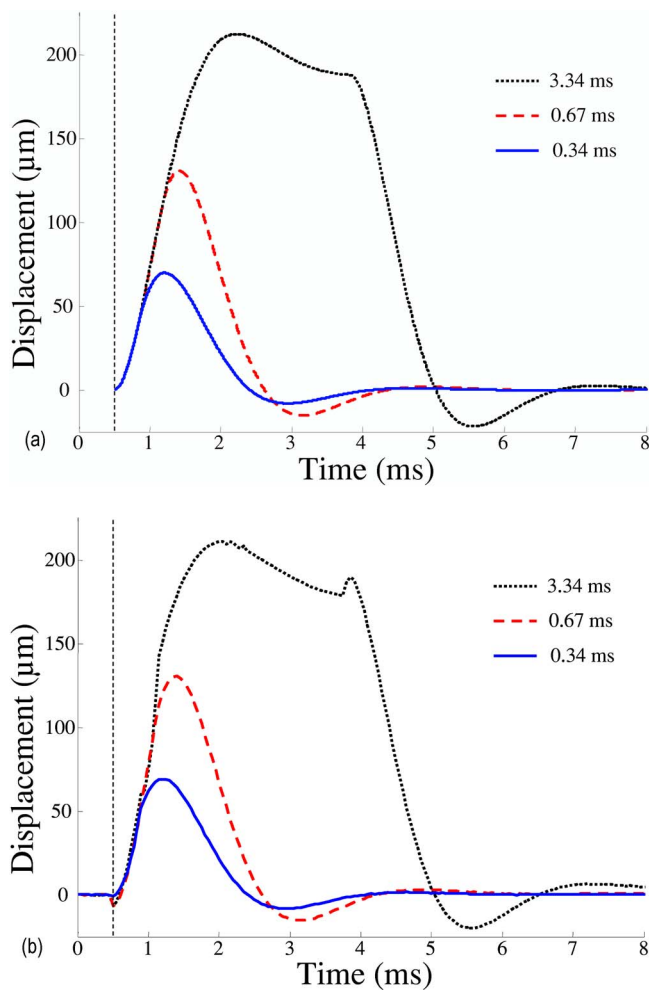
FIG. 4. (Color online) Displacements of the 1 mm diameter sphere (sphere B in Table I) in response to excitation pulses of different (0.34, 0.67, and 3.34 ms) durations. Shear elasticity of media was 1300 Pa. (a) Theory and (b) experiment.

FIG. 5. (Color online) Comparison of theoretically predicted and experimentally measured behavior of sphere D (see Table I) embedded into gelatin with 600 Pa shear modulus. In the theoretical model, the applied force was 2.6 mN. (a) Maximum displacement of the sphere vs excitation time. (b) Time needed to reach maximum displacement vs excitation time.

tween theory and experiment is nearly perfect although experimental estimates have a finite signal-to-noise ratio. The measurements are sometimes prone to signal processing artifacts due to electromagnetic interference noise introduced by the rf amplifier at the beginning and at the end of the pushing pulses, which results in signal decorrelation, as is evident in the inaccurate estimate at the end of the 3.34 ms long pushing pulse [Fig. 4(b)].

The maximum displacement $U(t_{max})$ and the time needed to reach maximum displacement t_{max} were analyzed. The dependences of these quantities on the duration t_0 of the excitation pulse are presented in Figs. 5(a) and 5(b) for theoretical and experimental data, respectively. The experiments were performed using a 3 mm glass sphere (sphere D in Table I). The time when the displacement reaches the saturation point t_{max}^0 is about 7 ms, and for $t_0 > t_{max}^0$ the time of maximum displacement is constant: $t_{max} = t_{max}^0$.

Theoretical and experimental results for spheres with various densities (spheres D, E, F, and G in Table I) are presented in Figs. 6(a) and Fig. 6(b). Because material properties of the spheres are different, the radiation force applied to spheres cannot be considered the same even though the spheres have the same size.[29] However, the magnitude of the

radiation force is only a scalar factor and, in linear approximation, it does not influence temporal behavior of the sphere motion.

Therefore, normalized displacements were used to compare experimental measurements and numerical estimations. All other parameters except density were the same. The time of maximum displacement increases with increased density of the sphere. In addition, the oscillations are more pronounced for denser spheres. Displacement of the lighter plastic sphere (sphere F in Table I) exhibits oscillations in neither the theoretical model nor in the experiment although the normalized density for this sphere is slightly more than 5/8. This is explained by a non-zero shear viscosity (0.1 Pa s) of the medium. Again, there is good agreement between experimental and theoretical data.

Figures 7(a) and 7(b) compare theoretical and experimental results obtained for a 3 mm light sphere (sphere F in Table I) embedded in a medium with changing elastic properties. Shear moduli were measured 990, 1350, 1550, and 1850 Pa at 20, 18, 14, and 10 °C temperatures of the phantom, correspondingly. Increase in elasticity of the surrounding material lead to decrease of both displacement magnitude
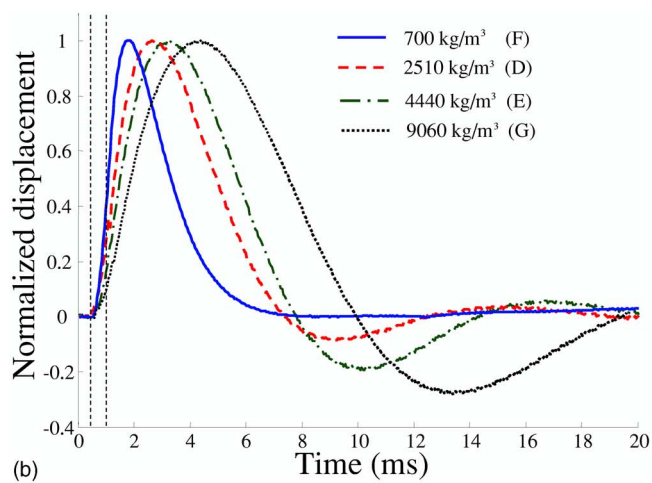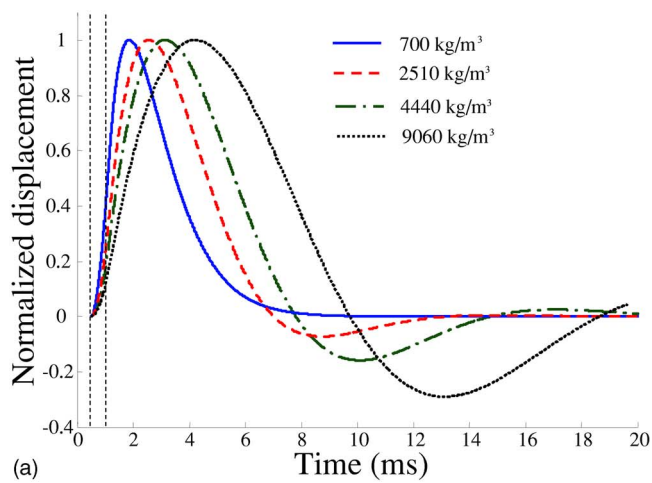
FIG. 6. (Color online) Normalized displacements of 3 mm diameter spheres (spheres D, E, F, and G) with different densities. Excitation time was 0.67 ms and shear elasticity of media was 600 Pa. (a) Theory and (b) experiment.

FIG. 7. (Color online) Displacements of the 3 mm diameter light sphere (sphere F) in a medium with different shear modulus. Shear moduli was measured 990, 1350, 1550, and 1850 Pa at 20, 18, 14, and 10 °C temperatures of the phantom, correspondingly. Excitation time was 0.67 ms. (a) Theory, applied force was 4.6 mN and (b) experiment.

and time needed for the displacement to reach the maximum. This is clearly demonstrated from theoretical data in Fig. 7(a). These results are in agreement with theoretical and experimental results obtained by other groups,[8,17,18,20] where the response to acoustical radiation force was investigated for a medium with different elasticity.

To investigate the influence of media elasticity on sphere behavior under various durations of radiation force, the theoretical and experimental studies were performed for two different temperatures (20 and 6 °C). The 3 mm diameter glass sphere (sphere E in Table I) was embedded in the phantom. The shear elastic moduli of the phantom for 20 and 6 °C were estimated as 650 and 3110 Pa, respectively. The theoretical and experimental results for 650 Pa shear elasticity are presented in Figs. 8(a) and 8(b), where the time dependences for different time of excitation are shown. Figures 9(a) and 9(b) present the same results for a sphere in the medium with 3110 Pa shear elasticity. These results show that the displacements are higher for soft material and the saturation of displacement starts earlier for harder material. For a medium with 650 Pa shear elasticity, the maximum displacement continues to increase even for the excitation time of 6 ms. For a medium with 3110 Pa shear elasticity,

however, the saturation point is already reached for acoustic radiation force pulses shorter than 4 ms. It should also be noted that even though the same phantom was used for both sets of the experiments, the estimated radiation forces were different: 1.65 mN for 650 Pa and 1.2 mN for 3110 Pa. This can be attributed to temperature-dependent acoustic properties of gelatin.[30] However, theoretical and experimental temporal characteristics of sphere motion are in good agreement and our results generally agree with results of Erpelding et al.[7] where the decrease of the maximum time for hard material was measured using a gas bubble as a target for acoustical radiation force. Figure 10 outlines the relationship between the time needed to reach the maximum displacement and the excitation time. Again, the two values of elasticity were considered.

Results shown in Figs. 7–9 demonstrate that the displacement of the sphere decreases when the elasticity of the medium increases. However, it is difficult to rely on the displacement amplitude alone to estimate the elasticity of the medium. In addition to shear elastic modulus, the displacement amplitude also depends on the magnitude of the radiation force applied to the sphere. However, the energy deliv-
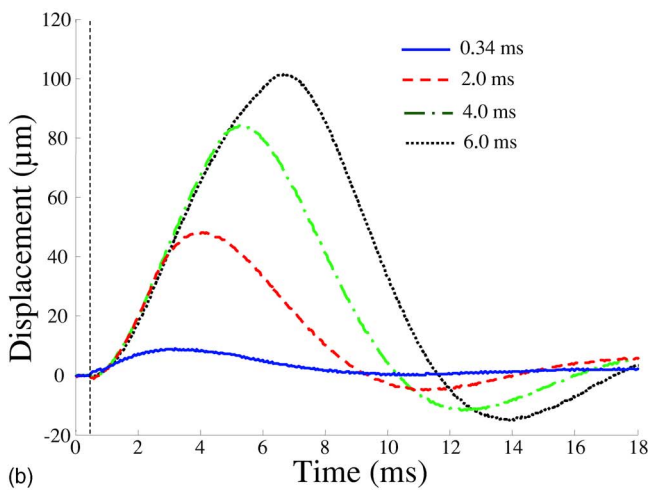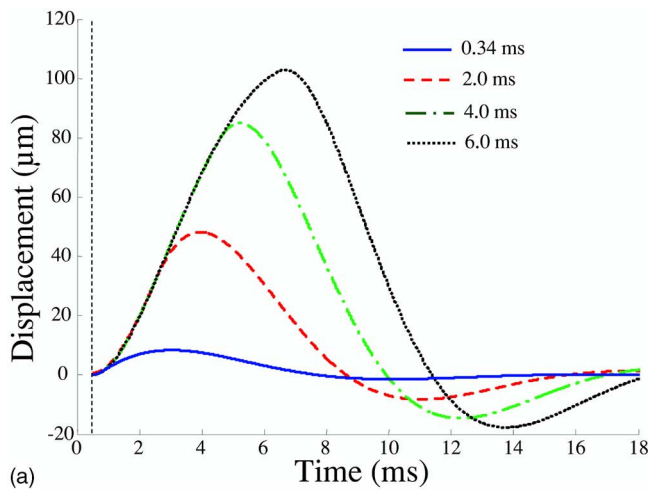
FIG. 8. (Color online) Displacements of the 3 mm diameter sphere (sphere E) embedded in the medium with shear elasticity 650 Pa (temperature 20 °C) in response to excitation pulses of different durations. (a) Theory, applied force was 1.65 mN and (b) experiment.
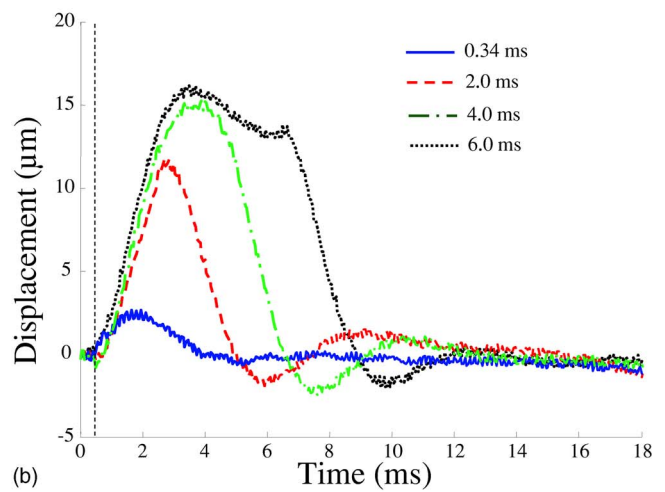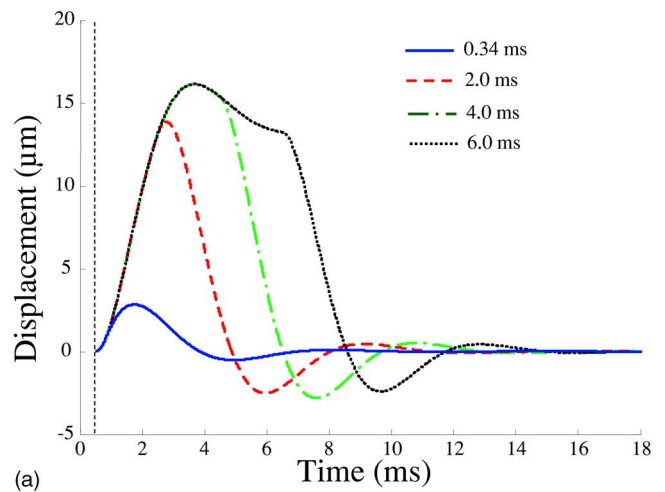
FIG. 9. (Color online) Displacements of the 3 mm diameter sphere (sphere E) embedded in the medium with shear elasticity 3110 Pa (temperature 6 °C) in response to excitation pulses of different durations. (a) Theory, applied force was 1.2 mN and (b) experiment.

ered to the tissue located in the focal spot of the pushing beam is unknown given the differences in attenuation of various tissues and the evaluation of radiation force magnitude presents a challenge.

On the contrary, the time characteristics of the sphere motion are independent of the radiation force amplitude because the force is only a scaling factor in the model describing the displacement of the sphere. Figure 11 illustrates this point: the displacements of a 3 mm diameter metal sphere (sphere G) are contrasted for radiation force pulses of different magnitude (the pre-amplified voltage applied to the pushing transducer is normalized to the maximum value). The duration of the pulses was set to 0.67 ms in all experiments. As is evident from Fig. 11, the amplitude of the displacement is proportional to the amplitude of radiation force, but temporal characteristics of the displacement (e.g., time needed to reach the maximum of the displacement or time needed for the sphere to come back to the point of the origin) remain the same. These results suggest that elastic properties of the tissue can be estimated from temporal rather than spatial characteristics of the sphere motion.

In general, the motion of the sphere in tissue is related to both elasticity and viscosity of the medium. The theoretical
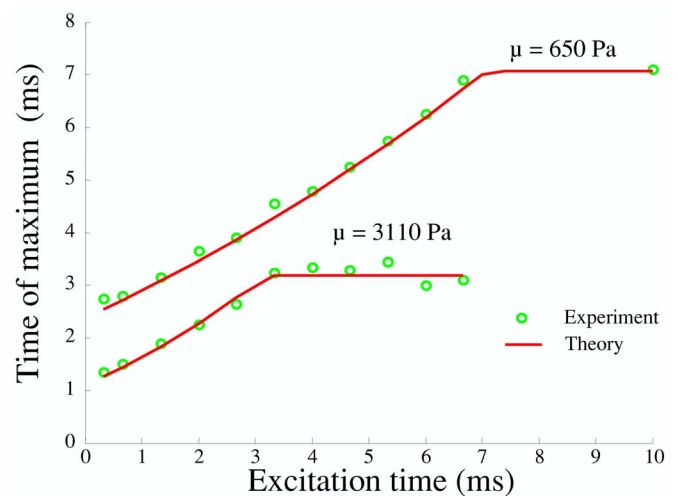
FIG. 10. (Color online) Comparison of theoretically predicted and experimentally measured dependences of time of maximum displacement on excitation time for the 3 mm sphere (sphere E) and for two different shear elasticity coefficients: $\mu = 650$ Pa (temperature 20 °C) and $\mu = 3110$ Pa (temperature 6 °C).

Aglyamov *et al.*: Solid sphere motion in viscoelastic medium
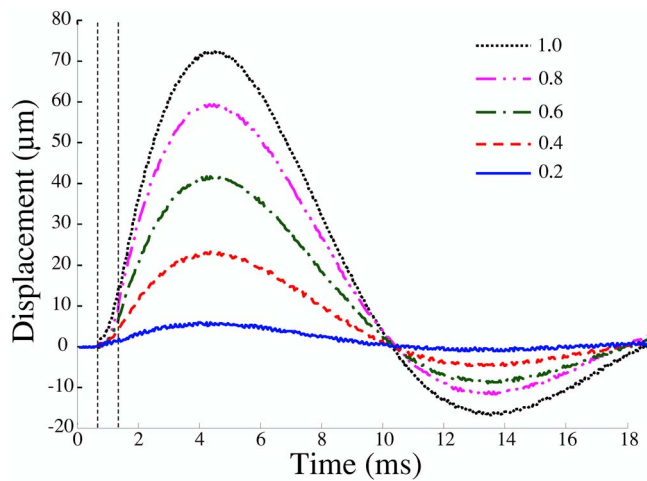
FIG. 11. (Color online) Displacements of the 3 mm diameter metal sphere (sphere G) in response to 0.67 ms long pulses of various amplitudes (normalized to a maximum value). Shear elasticity of media was 600 Pa. Note temporal similarity of motion.

analysis and numerical studies for solid sphere and gas bubble dynamics in a viscoelastic medium are presented elsewhere.[24] The displacement of a solid sphere or gas bubble in the medium with no viscosity has an oscillatory character while viscosity eliminates oscillations in the displacements and reduces the maximum value of the displacement amplitude. The formalism developed here suggests that both shear elasticity and shear viscosity may be evaluated from the response of the sphere to the radiation force excitation. In this work, however, influence of shear viscosity on sphere behavior is not considered for two reasons. First, there is not a standard method to measure independently the shear viscosity of tissues and tissue-like materials. Gold standard methods exist only for liquids, but these methods cannot be directly used for solids. Hence, we cannot reliably verify our estimation of shear viscosity modulus. Second, our model-based estimation of the shear viscosity for gel phantoms used in the acoustical radiation force experiments resulted in 0.1 Pa s. As the elastic modulus was modified by changing the temperature of the phantom, the viscosity was also expected to change. However, the slight variations of the shear viscosity because of temperature variations did not significantly affect the results. In accordance with our theoretical analysis, the influence of shear viscosity in this range is negligible and elasticity of the medium largely defines the behavior of the sphere. Nevertheless, our shear viscosity estimation is in agreement with the results obtained by several research groups.[19,21] For example, the shear viscosity estimation based on the measurement of shear wave amplitude attenuation gave a value of 0.1 Pa s for a 4% gelatin phantom.[21] Others estimated the shear viscosity of gelatin phantoms using two independent methods: measuring shear wave speed dispersion and measuring resonance frequencies of an embedded solid sphere.[19] The shear viscosity for 12% gelatin phantoms was found in the range of 0.2–0.4 Pa s. Taking into account that we used a 3% by weight concentration of gelatin, the viscosity estimation in 0.1 Pa s appears to be reasonable. However, additional studies are needed to

fully investigate the influence of shear viscosity on oscillatory and translational motion of the sphere in a viscoelastic medium.

The results of our study can be applied in many biomedical and clinical fields. For example, the model developed here could possibly assist laser-based diagnosis and microsurgery in ophthalmology. The microbubbles created during laser-tissue interaction in the cornea, lens and vitreous humor, are used as part of surgery procedures.[31,32] In each case, using acoustic radiation force applied to laser induced microbubbles, the mechanical properties of the eye components could be estimated to improve diagnosis or therapeutic strategy. There are no fundamental differences between solutions describing dynamic behavior of a solid sphere and an air bubble in a viscoelastic medium. The difference is only in the boundary conditions on the surface of the spherical object.[24] Therefore, air bubbles can be used as targets for the acoustical radiation force measurements[6,7] thus allowing remote estimation of tissue elasticity of the cornea, lens and vitreous humor. However, further experiments are needed to validate the theory of gas microbubbles behavior in viscoelastic tissue.

V. CONCLUSIONS

The results of this study demonstrate that the theoretical model predicts the motion of the solid spheres in viscoelastic medium in response to the acoustic radiation force. Our investigation of the sphere response to the acoustic radiation force for various sphere sizes and densities in media with different elastic properties shows that there is a good agreement between theoretical predictions and experimental measurements.

ACKNOWLEDGMENTS

[1]Y. C. Fung, *Biomechanics. Mechanical Properties of Living Tissues*, 2nd ed. (Springer, New York, 1993).

[2]J. Ophir, S. K. Alam, B. Garra, F. Kallel, E. Konofagou, T. Krouskop, and T. Varghese, "Elastography: Ultrasonic estimation and imaging of the elastic properties of tissues," Proc. Inst. Mech. Eng., Part H: J. Eng. Med. **213**, 203–233 (1999).

[3]S. R. Aglyamov and A. R. Skovoroda, "Mechanical properties of soft biological tissues," Biophycics **45**, 1103–1111 (2000).

[4]A. P. Sarvazyan, "Elastic properties of soft tissues," in *Handbook of Elastic Properties of Solids, Liquids and Gases*, edited by M. Levy, H. Bass, and R. Stern (Academic, New York, 2001), Vol. **3**, pp. 107–127.

[5]L. A. Negron, F. Viola, E. P. Black, C. A. Toth, and W. F. Walker, "Development and characterization of a vitreous mimicking material for radiation force imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1543–1551 (2002).

[6]T. N. Erpelding, R. C. Booi, K. W. Hollman, and M. O'Donnell, "Measuring tissue elastic properties using acoustic radiation force on laser-generated bubbles," Proc.-IEEE Ultrason. Symp. **1**, 554–557 (2003).

[7]T. N. Erpelding, K. W. Hollman, and M. O'Donnell, "Bubble-based acoustic radiation force elasticity imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 971–979 (2005).

[8]K. Nightingale, M. S. Soo, R. Nightingale, and G. Trahey, "Acoustic radiation force impulse imaging: In vivo demonstration of clinical feasibility," Ultrasound Med. Biol. **28**, 227–235 (2002).

[9]M. Fatemi and J. F. Greenleaf, "Ultrasound-stimulated vibro-acoustic spectrography," Science **280**, 82–85 (1998).

[10]F. L. Lizzi, R. Muratore, C. X. Deng, J. A. Ketterling, S. K. Alam, S.

Mikaelian, and A. Kalisz, "Radiation-force technique to monitor lesions during ultrasonic therapy," Ultrasound Med. Biol. **29**, 1593–1605 (2003).

[11] K. R. Nightingale, R. W. Nightingale, D. L. Stutz, and G. E. Trahey, "Acoustic radiation force impulse imaging of in vivo vastus medialis muscle under varying isometric load," Ultrason. Imaging **24**, 100–108 (2002).

[12] F. G. Mitri, P. Trompette, and J. Y. Chapelon, "Improving the use of vibro-acoustography for brachytherapy metal seed imaging: A feasibility study," IEEE Trans. Med. Imaging **23**, 1–6 (2004).

[13] T. Sugimoto, S. Ueha, and K. Itoh, "Tissue hardness measurement using the radiation force of focused ultrasound," Proc.-IEEE Ultrason. Symp. **3**, 1377–1380 (1990).

[14] M. Fatemi and J. F. Greenleaf, "Application of radiation force in noncontact measurement of the elastic parameters," Ultrason. Imaging **21**, 147–154 (1999).

[15] W. F. Walker, "Internal deformation of a uniform elastic solid by acoustic radiation force," J. Acoust. Soc. Am. **105**, 2508–2518 (1999).

[16] K. R. Nightingale, M. L. Palmeri, R. W. Nightingale, and G. E. Trahey, "On the feasibility of remote palpation using acoustic radiation force," J. Acoust. Soc. Am. **110**, 625–634 (2001).

[17] A. R. Skovoroda and A. P. Sarvazyan, "Determination of viscoelastic shear characteristics of a medium from its response to focused ultrasonic loading," Biophycics **44**, 325–329 (1999).

[18] S. Calle, J. P. Remenieras, O. B. Matar, M. E. Hachemi, and F. Patat, "Temporal analysis of tissue displacement induced by a transient ultrasound radiation force," J. Acoust. Soc. Am. **118**, 2829–2840 (2005).

[19] S. Chen, M. Fatemi, and J. F. Greenleaf, "Quantifying elasticity and viscosity from measurement of shear wave speed dispersion," J. Acoust. Soc. Am. **115**, 2781–2785 (2004).

[20] A. P. Sarvazyan, O. V. Rudenko, S. D. Swanson, J. B. Fowlkes, and S. Y. Emelianov, "Shear wave elasticity imaging: A new ultrasonic technology of medical diagnostics," Ultrasound Med. Biol. **24**, 1419–1435 (1998).

[21] J. Bercoff, M. Tanter, M. Muller, and M. Fink, "The role of viscosity in the impulse diffraction field of elastic waves induced by the acoustic radiation force," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 1523–1536 (2004).

[22] K. R. Nightingale, L. Zhai, J. J. Dahl, K. D. Frinkley, and M. L. Palmeri, "Shear wave velocity estimation using acoustic radiation force impulsive excitation in liver in vivo," Proc.-IEEE Ultrason. Symp. 1156–1160 (2006).

[23] S. Chen, M. Fatemi, and J. F. Greenleaf, "Remote measurement of material properties from radiation force induced vibration of an embedded sphere," J. Acoust. Soc. Am. **112**, 884–889 (2002).

[24] Y. A. Ilinskii, G. D. Meegan, E. A. Zabolotskaya, and S. Y. Emelianov, "Gas bubble and solid sphere motion in elastic media in response to acoustic radiation force," J. Acoust. Soc. Am. **117**, 2338–2346 (2005).

[25] A. P. Sarvazyan, "Low frequency acoustic characteristics of biological tissues," Mech. Polymers **4**, 691–695 (1975).

[26] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed. (Pergamon, New York, 1987).

[27] H. L. Oestreicher, "Field and impedance of an oscillating sphere in a viscoelastic medium with an application to biophysics," J. Acoust. Soc. Am. **23**, 707–714 (1951).

[28] M. A. Lubinski, S. Y. Emelianov, and M. O'Donnell, "Cross-correlation speckle tracking techniques for ultrasound elasticity imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 82–96 (1999).

[29] T. Hasegawa and K. Yosioka, "Acoustic-radiation force on a solid elastic sphere," J. Acoust. Soc. Am. **46**, 1139–1143 (1969).

[30] E. L. Madsen, J. A. Zagzebski, R. A. Banjavie, and R. E. Jutila, "Tissue mimicking materials for ultrasound phantoms," Med. Phys. **5**, 391–394 (1978).

[31] A. Vogel, M. R. C. Capon, M. N. Asiyo-Vogel, and R. Birngruber, "Intraocular photodisruption with picosecond and nanosecond laser pulses: Tissue effects in cornea, lens, and retina," Invest. Ophthalmol. Visual Sci. **35**, 3032–3044 (1994).

[32] H. Lubatschowski, G. Maatz, A. Heisterkamp, U. Hetzel, W. Drommer, H. Welling, and W. Ertmer, "Application of ultrashort laser pulses for intrastromal refractive surgery," Graefe's Arch. Clin. Exp. Ophthalmol. **238**, 33–39 (2000).

Prediction of noise levels and annoyance from aircraft run-ups at Vancouver International Airport

Katrina Scherebnyj and Murray Hodgson[a)]
Acoustics and Noise Research Group, MECH-SOEH, 3rd Floor, 2206 East Mall, Vancouver,
British Columbia V6T 1Z3, Canada

Annoyance complaints resulting from engine run-ups have been increasing at Vancouver International Airport for several years. To assist the Airport in managing run-up noise levels, a prediction tool based on a Green's function parabolic equation (GFPE) model has been consolidated, evaluated, and applied. It was extended to include more realistic atmospheric and ground input parameters. Measurements were made of the noise-radiation characteristics of a CRJ200 jet aircraft. The GFPE model was validated by comparing predictions with results in the literature. A sensitivity analysis showed that predicted levels are relatively insensitive to small variations in geometry and ground impedance, but relatively sensitive to variations in wind speed, atmosphere type, and aircraft heading and power setting. Predicted noise levels were compared with levels measured at noise monitoring terminals. For the four cases for which all input information was available, agreement was within 10 dBA. For events for which some information had to be estimated, predictions were within 20 dBA. The predicted annoyance corresponding to the run-up events considered ranged from 1.8% to 9.5% of people awoken, suggesting that noise complaints can be expected. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2769988]

I. INTRODUCTION

Vancouver International Airport Authority (YVRAA), the body that manages Vancouver International Airport, receives hundreds of noise complaints each year. A significant portion results from engine run-ups by jet or propeller aircraft. A run-up—the testing of a stationary aircraft's engines—is a routine procedure that occurs following aircraft maintenance. There are certain parameters of engine run-ups, such as the location and heading of the aircraft, that can affect community noise levels, and that the airport can manage. This could be done more effectively using a prediction tool that could predict noise levels and annoyance in communities surrounding the airport that result from aircraft run-up noise.

The propagation of noise outdoors is a complex phenomenon. As sound propagates, it can be reflected, refracted, attenuated, and amplified.[1,2] Many factors influence outdoor sound propagation, including complex atmospheric and ground conditions. The commonly assumed hemi-free-field conditions over a reflective ground are not realistic. The work reported here extends previous work, which used a Green's function parabolic equation (GFPE) model in simplified atmospheric and ground conditions.[3]

The objective of the present work was to use more realistic environmental conditions—such as realistic wind-speed and temperature profiles, and mixed ground impedance—to predict noise levels and associated community reaction, in residential areas surrounding the Vancouver International Airport, resulting from propeller- and jet-aircraft engine run-

ups, to help YVR minimize community annoyance for given atmospheric conditions and airline requirements. Full details of the work described here are published elsewhere.[4]

II. RUN-UP MEASUREMENTS

Accurate noise prediction requires accurate aircraft sound-source data. Data for run-ups by two propeller aircraft—a DeHavilland Dash-8 and a Beechcraft 1900— were available from previous work.[3,5] It was of interest here to obtain data for a jet aircraft. The purpose of these measurements was to determine the energetic and directional radiation characteristics (spectra and directivities) of a jet aircraft that operates at Vancouver International Airport, and whether it is similar to or different from those for the propeller aircraft measured previously. Of course, the CRJ200 is smaller than many jet aircraft; its noise radiation does not necessarily represent that of other jet aircraft.

Noise measurements during the run-up of a CRJ200 were made in conjunction with a taxiing exercise late one evening in August 2005. The wind speed was 4 m/s and the temperature was 17 °C (at 10 and 6 m above ground level, respectively). Equivalent-continuous sound-pressure levels (L_{eq}) were measured in third-octave bands from 25 to 8000 Hz at two heights (ground level and 1.4 m above the ground), at 15 locations around the aircraft, 40 m from the center of the source, as illustrated in Fig. 1, and at three power settings (idle, 50% power, and full power). The run-up took place in a large, open area, over concrete. Measurement positions located at the rear of the aircraft were not measured, for safety reasons. Third-octave-band background-noise levels were at least 10 dB below the signal levels in all cases (except at a few positions, for some frequencies below

---
a)Electronic mail: hodgson@mech.ubc.ca
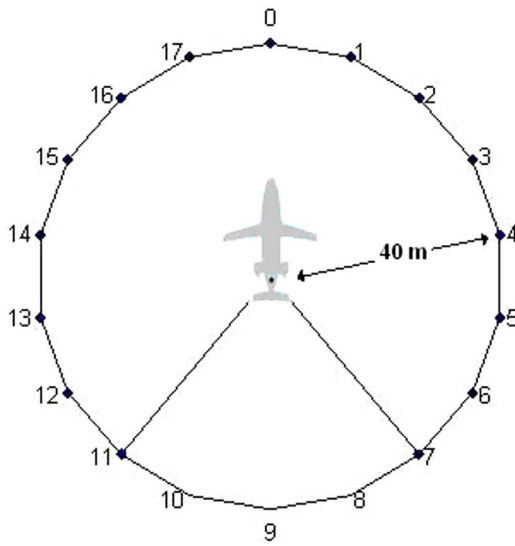
FIG. 1. (Color online) Measurement locations for the jet-aircraft noise measurements.

FIG. 2. Measured CRJ200 unweighted, third-octave band L_{eq} spectra at Positions (a) 0, (b) 3, and (c) 7 (see Fig. 1), at two heights, for three engine-power settings. Full power: (—) 1.4 m, (—) ground level; 50% power: (– – –) 1.4 m, (- - -) ground level; idle ($\cdots$) 1.4 m, ($\cdots$) ground level.

50 Hz, for which the difference was between 5 and 10 dB). The unweighted L_{eq} spectra at positions in front, to one side, and toward the back, of the aircraft are shown in Fig. 2. Results for Position 2 were not obtained due to equipment malfunction. The levels measured at ground level tended to be higher than those at 1.4 m above the ground, as expected, since destructive interference can result in lower sound-pressure levels at receiver positions above the ground. Two major distinctions between the CRJ200 and the propeller-aircraft noise characteristics are in the shapes of the sound-pressure-level spectra, and their magnitudes. The propeller aircraft displayed prominent tonal components at lower frequencies (e.g., at the blade-passage frequency and its harmonics), particularly at higher-power engine settings, unlike the more broadband spectra of the CRJ200. Furthermore, the levels generated by the CRJ200 tended to be higher—for example, 10–30 dB higher than the Dash-8.

Figure 3 shows the measured total, unweighted L_{eq} directivity results for the CRJ200. Results indicate that the radiation from this jet aircraft was quite axisymmetric; levels on the two sides of the aircraft were usually within a couple of decibels of one another—the average difference was around 1 dB. The directivity of the source varies by 5 dB or more at different engine-power settings. These results suggest that levels behind the aircraft, where measurements were not possible, might be fairly uniform and similar to those at the rear-most positions that were measured. Germain et al.[5] measured levels behind a propeller aircraft, finding variations of total, unweighted level of up to 4–7 dB.

III. GFPE MODEL

A. Model description

There are several outdoor-sound-propagation prediction models available.[6] More common models include the generalized fast field program (FFP), the parabolic equation (PE) method, and ray-tracing. Since FFP models are restricted to environments with layered atmospheres and homogeneous ground surfaces, and since ray-tracing programs become extremely computationally expensive for the long ranges involved outdoors and must deal with the appearance of "caustics," the PE method was chosen for use here. PE models can incorporate mixed ground impedance and sound-speed profiles. There are several ways to solve the parabolic equation numerically—the Crank-Nicholson parabolic equation (CNPE) method and the GFPE method are two common methods. The horizontal and vertical step sizes for the CNPE are limited to a maximum of $\lambda/10$ (λ is wavelength). In the GFPE, the vertical step size has the same limitation, but the horizontal step size can be between 10λ and 100λ, leading to a major reduction in computation time, particularly at higher frequencies (X. Di, private communication).[7]
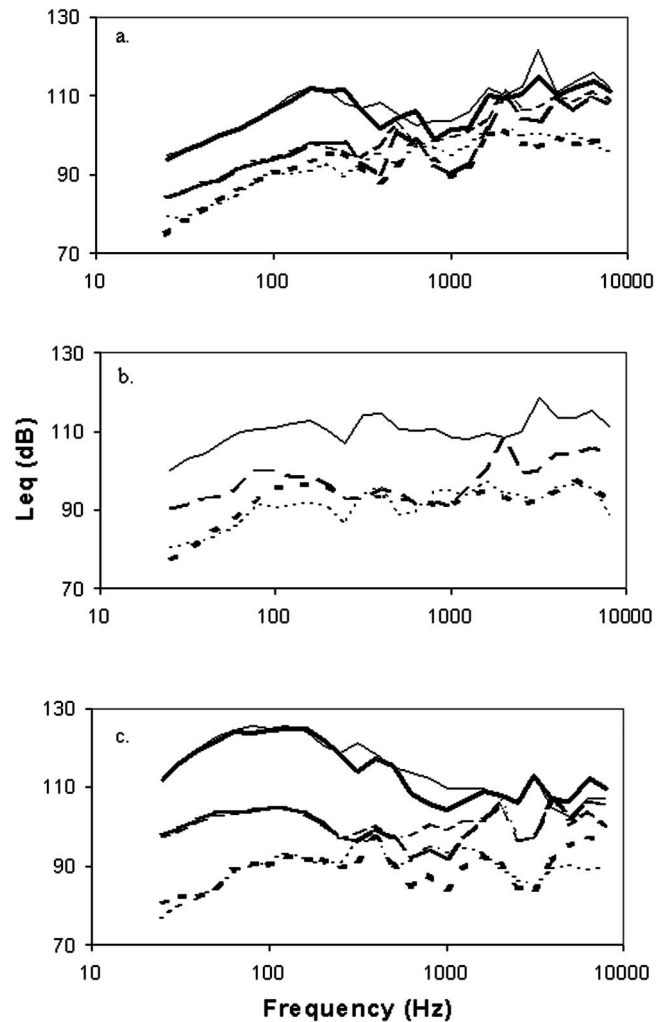
The theory behind the GFPE model is based on the Helmholtz equation, into which a new, simplified quantity that has cylindrical spreading removed is substituted for sound pressure. It is then assumed that the majority of sound propagates in the forward direction and that the wave number is constant across a step Δr. Integrating across the step subsequently solves the Helmholtz equation. The GFPE model solves for sound pressure on a two-dimensional (r, z)
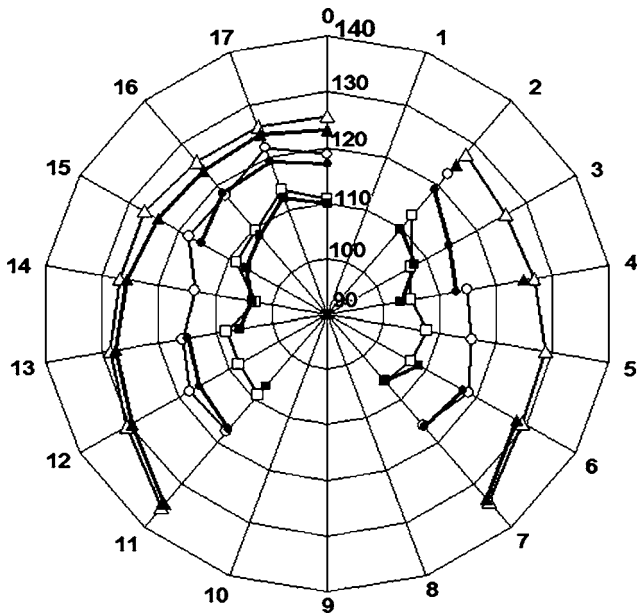
FIG. 3. Measured CRJ200 total, unweighted L_{eq} directivity at 40 m at positions shown in Fig. 1, at two heights, for three engine-power settings. Full power: (▲) 1.4 m, (△) ground level; 50% power: (●) 1.4 m, (○) ground level; idle (■) 1.4 m, (□) ground level.

grid. The horizontal dimension is the source-receiver distance, divided by steps Δr; the vertical dimension is divided into steps Δz. The domain of r is from the first step $(0 + \Delta r)$ to the receiver position. The sound field at $r=0$ is represented by a Gaussian function. The domain of z is from the ground level, where the reflection coefficient determines how much sound is reflected, to an upper absorbing layer, set by the user. Rather than have the upper-atmosphere limit end abruptly, an absorbing layer that is at least 30λ thick exists above the upper limit of the atmosphere through which sound propagates, to prevent sound from reflecting unrealistically back down into the atmosphere. Unfortunately, direct integration across the steps Δr is not possible; instead, the expression is solved using Green's function, and either spectral representation or using Rayleigh's integral. This process is repeated, iteratively marching through the solution in steps of Δr, until the receiver sound-pressure level is predicted.[6,8]

B. Modifications

A two-dimensional GFPE model, originally developed by Gilbert and Di,[8] was available from previous work,[3] and was extended. The model calculates the complex sound pressure at a single frequency as a function of the distance over homogeneous ground of arbitrary acoustical impedance, with user-defined vertical sound-speed gradients.[8] Improvements implemented to achieve the objectives of the present work included accounting for nonhomogeneous ground (variations in ground type/impedance with distance), the prediction of sound-pressure levels in third-octave bands, implementing a frequency-dependent step size, accounting for air absorption, the calculation of total A-weighted levels, the automatic calculation of the sound-speed profile given the temperature and wind conditions, the estimation of sound-exposure level (SEL), and the prediction of the annoyance associated with
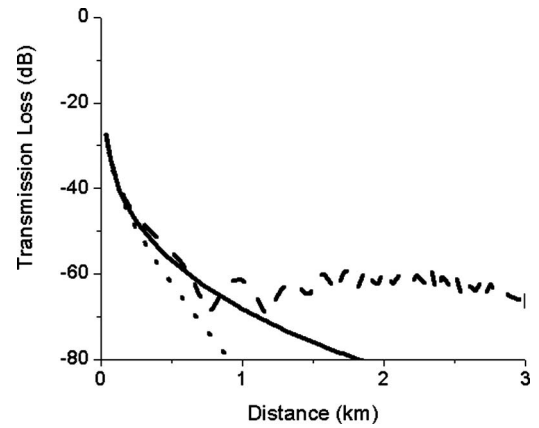
FIG. 4. GFPE predictions of the variation of transmission loss with distance at 100 Hz in (—) homogeneous, (···) upward-refracting, and (- - -) downward-refracting atmospheres (see the text).

the noise. Six predefined, finite ground impedances were programmed into the model, based on the grounds that were considered typical of the areas at and around YVR (dry grass, wet grass, dry sand, wet sand, snow, and an acoustically hard surface corresponding to concrete, water, or asphalt). The corresponding impedances were calculated according to Attenborough's two-parameter model.[9]

C. Validation

The modifications made to the GFPE model were evaluated in comparison with results in the literature. Levels were predicted at various frequencies and over various distance ranges. Data from the Attenborough *et al.*[10] benchmark paper were used in the cases of homogeneous and upward- and downward-refracting atmospheres. Data from Gauvreau *et al.*[11] and Daigle *et al.*[12] were used in the case of mixed-ground impedance. Very good agreement (typically within 2 to 3 dB) was obtained in all cases (see Ref. 4).

Two representative validation results are shown in Figs. 4 and 5. First, the case of homogeneous ground impedance and different atmospheric conditions was considered. The input-parameter values were as follows: $h_{source}=5.0$ m,
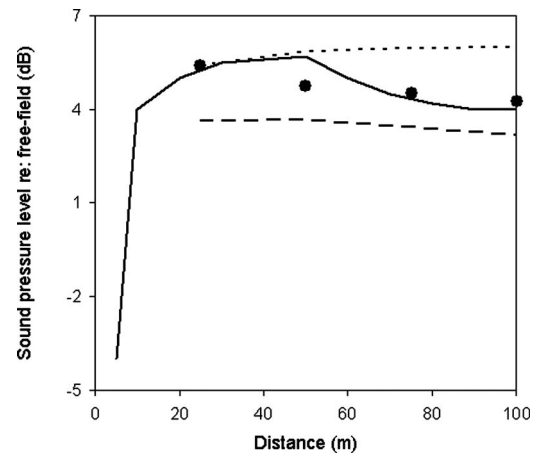
FIG. 5. Predicted variation with distance of sound-pressure level relative to free field for a mixed ground-impedance configuration (see the text): (●) GFPE; (—) CNPE (Ref. 11). Also shown for reference are the GFPE predictions for completely (···) "hard" and (- - -) "soft" grounds.

$h_{\text{receiver}} = 1.0$ m, normalized ground impedance $= (12.81, 11.62)$, $\Delta z = \lambda/5$, $h_{\text{vertical}} = 250$ m, $c = 343$ m/s, FFT length $= 16\,384$, surface-wave integration number $= 250$. Figure 4 shows GFPE predictions of the variation of transmission loss (sound-pressure level relative to that at 1 m in a free-field) with distance up to 3 km at 100 Hz in homogeneous, upward-refracting (sound-speed gradient $= +0.1$ s^{-1}) and downward-refracting (sound-speed gradient $= -0.1$ s^{-1}) atmospheres. Levels are very similar to those presented in Ref. 10. Figure 5 compares levels predicted by the GFPE for a configuration with mixed ground impedance with CNPE values estimated from Ref. 11. There is a ground impedance discontinuity 50 m from the source; the ground nearest the source is "hard" (flow resistivity $= 2 \times 10^5$ kPa s m^{-2}), that farthest from the source "soft" (flow resistivity $= 2 \times 10^2$ kPa s m^{-2}). The other input-parameter values are as follows: frequency $= 160$ Hz, $h_{\text{source}} = 1.5$ m, $h_{\text{receiver}} = 1.8$ m, $\Delta z = \lambda/10$, $h_{\text{vertical}} = 100$ m, $c = 340$ m/s, FFT length $= 16\,384$, surface-wave integration number $= 100$. Also shown for reference are the GFPE predictions for uniformly hard and soft grounds. The GFPE predictions are credible and agree well with the CNPE prediction. Results were equally as good at other frequencies (see Ref. 4).

Modifications were also made to include turbulence in the GFPE model. Predictions gave results similar to those of Gilbert[13] (see Ref. 4). However, in preliminary prediction work, the effects of turbulence were not found to contribute significantly in homogeneous, downward-refracting, and in very weak upward-refracting (e.g., with a decrease of 1 m/s over 200 m), atmospheres. These conditions were typical of the run-up events considered in this work; thus, turbulence was not considered further.

IV. PREDICTION

A. Atmospheric assumptions

There are three general states that exist for a static atmosphere: stable, unstable, and neutral. In a stable atmosphere, the temperature increases with height. The shape of the temperature-profile increase is expected to be parabolic (D. Steyn, private communication). An unstable atmosphere is one for which the temperature decreases with height at a rate greater than 0.0098 °C/m (the dry adiabatic lapse rate of the atmosphere). Instability indicates that the vertical movement of air packets is not restricted. The temperature profile in a neutral atmosphere decreases at a rate of 0.0098 °C/m.

The temperatures at two heights—1.6 and 6 m—were available for the times of the run-up events considered in this work. Unstable conditions at night are uncommon (D. Steyn, private communication). Thus, if the change in temperature was positive, stable conditions were applied; if the change in temperature was negative, neutral conditions were assumed. To approximate stable profiles, the temperature at 6 m was input and a parabola was fit to this value, and to the values at 0 and 200 m, which were assumed to be 2 °C lower and 6 °C higher, respectively, than the temperature at 6 m. The temperature profile for neutral conditions was set to decrease at a rate of 0.0098 °C/m. Temperatures in the first 10 m
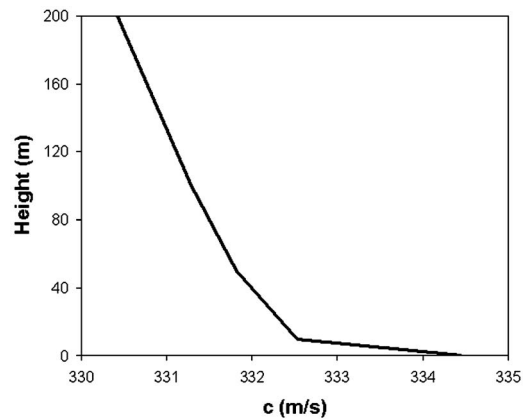
FIG. 6. Typical sound-speed profile used in GFPE predictions (details in the text).

above the ground do not follow a simple curve; this part of the profile is almost impossible to estimate without making measurements at several heights, adding an additional element of uncertainty to the assumptions made in prediction.

For both atmospheric states, the wind-profile power law was used to calculate the wind-speed profile (the variation of wind speed with height):

$$\frac{u(z)}{u(10)} = \left(\frac{z}{10} \right)^{1/7}, \tag{1}$$

in which u is wind speed and z is height above the ground. The wind vector was then projected onto the source-receiver direction and added to the sound speed. It was suspected that mixing due to winds greater than 5 m/s at 10 m would prevent the above-described stable and neutral atmospheric states from forming. Run-up events that took place during such times were therefore separated from those with lower winds in the analysis in Sec. IV C.

The sound-speed profile for a "typical" night-time run-up (18 April 2005; neutral conditions, temperature of 6 °C at 6 m above the ground, wind speed of 2.5 m/s at 10 m above the ground, incident at an angle of approximately 42° to the source-receiver direction) is illustrated in Fig. 6. In this particular case, the sound would be refracted upward as it propagates from the source to the receiver. The curve is not perfectly smooth, as the GFPE model integrates the sound-speed profile from a set of points calculated by Eq. (1).

B. Sensitivity analysis

To investigate the dependence of predicted noise levels on the input parameters of the GFPE model and, therefore, the expected influence of uncertainties in the input-parameter values, a simple configuration was chosen as a reference, and predictions for other cases were compared with levels predicted for it. The prediction input parameters were varied slightly, one at a time. A-weighted levels from 25 to 2000 Hz were predicted and compared. In one variation, the frequency range was also extended, for comparison with the reference spectral range. The parameter values for the reference case were as follows:
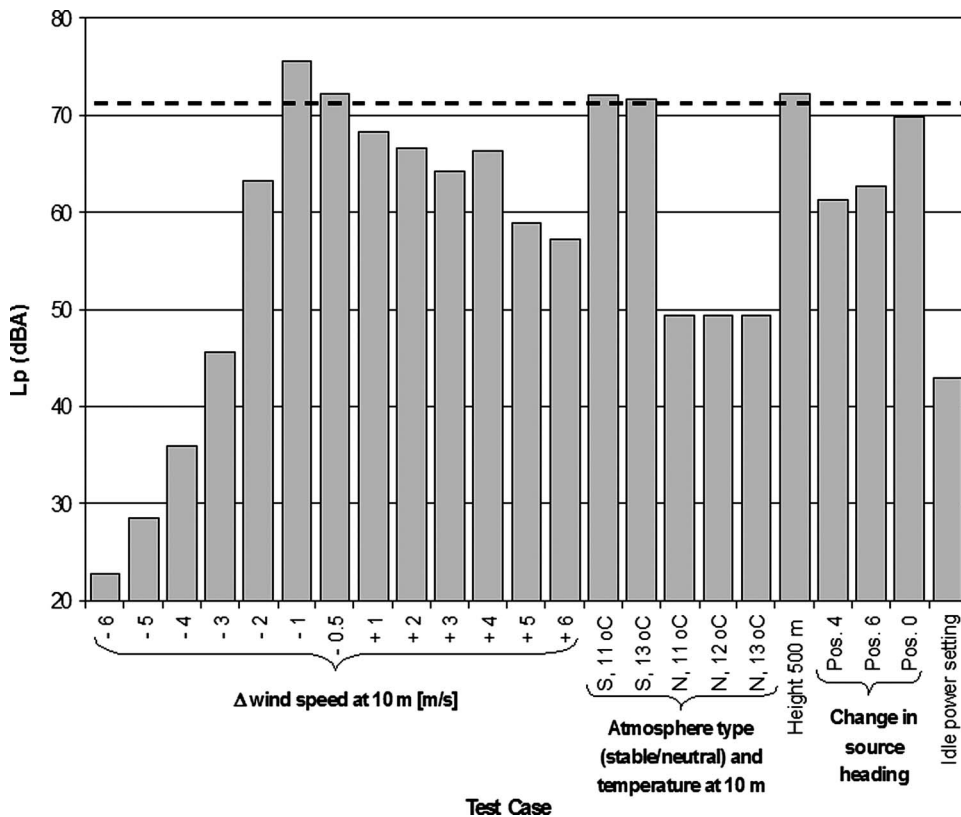
FIG. 7. Comparison of predicted total, A-weighted sound-pressure levels for the reference test case (dashed line at 71 dBA) with those for test cases with input-parameter variations in wind speed, temperature, height of propagation, and aircraft details.

(1) Source-receiver distance of 2 km;

(2) One ground-impedance transition point midway between the source and the receiver (at 1 km), separating "concrete" nearest the source from "dry, loamy grass" farthest from the source;

(3) Dash-8 propeller-aircraft source, running-up at full power; source height 3 m above the ground;

(4) Receiver height 6 m above the ground (a typical noise monitoring terminal microphone height); and

(5) Stable atmosphere, with no wind, and a temperature of 12 °C at 6 m.

Geometric variations included changes in the total source-receiver distance (calculated by adding or removing 10 m from either the grass or concrete section), shifting the transition point by 10 m, as well as changes in the source height (± 1 m) and receiver height ($\pm 0.1, 0.5$ m). The magnitudes of these changes were chosen to represent the uncertainties involved in estimating the input values. All of the resulting predictions were within 2 dB of the reference case.

Variations relating to the atmosphere included changes in the wind speed used in the wind-profile power law, changes in the temperatures at the reference heights, and switching from stable to neutral conditions. Atmospheric-variation predictions, along with predictions for a doubling of the numerical-grid height for which the sound-pressure levels are calculated (as mentioned in Sec. III A), and changes in the aircraft engine-power setting and heading, are presented in Fig. 7.

By way of these predictions it was found that, in general in the current application, the sensitivity of the GFPE model with respect to variations in most input parameters is rather

low: realistic variations do not yield very different results. The exceptions to this are the cases of wind speed, atmosphere type, and aircraft engine-power setting and heading. The values of the latter two factors are well known and, as long as communication between the aircraft operator and the airport is clear, this information should be readily accessible. This is, however, not the case for wind profiles, and for the state of the atmosphere: these will never be known accurately unless they are measured. Their dynamic nature makes the measurements difficult.

Temperature and wind profiles can strongly influence the sound-speed gradient, and thus the manner in which sound will be refracted as it propagates. In instances of very strong upward refraction where shadow zones are created, such as the extreme case with wind designated "−6 m/s@10 m," very large differences, such as the decrease of nearly 50 dB in Fig. 7, can occur. This is not to say that the outdoor sound-pressure level will actually be 23 dBA; background levels will invariably be higher, masking the contribution from the engine run-up. Moreover, turbulence will reduce the high attenuation at large distances.

C. Test cases

Run-ups that occurred at the Vancouver International Airport between January 2005 and June 2006 were drawn upon as the basis for comparisons between predicted and measured noise levels, to investigate prediction accuracy in realistic situations. In order for a run-up to be considered, someone in the community, whose address or approximate location was available, was required to have reported a noise complaint corresponding to the approximate time of the en-
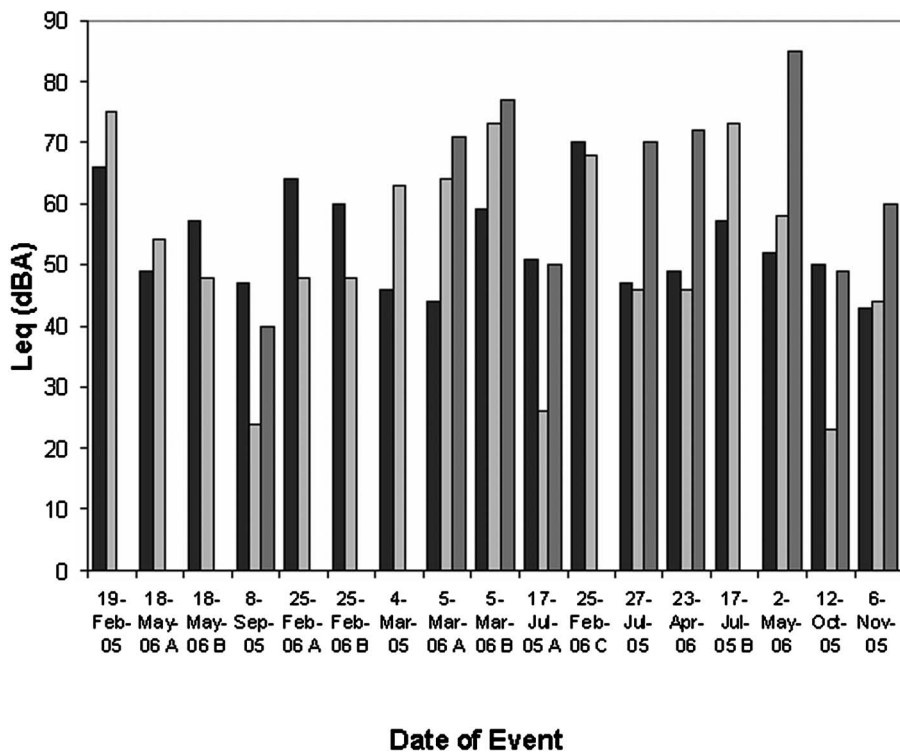
FIG. 8. Summary of GFPE-predicted total, A-weighted run-up noise levels and measured equivalent-continuous NMT levels: (closed square) measured; (gray square) "low" prediction; and (dark gray square) "high" prediction, where applicable.

gine run-up. Only night-time run-up events were considered. Seventeen such events were identified, and predicted sound-pressure levels were compared with the corresponding levels measured at noise monitoring terminals (NMTs); the annoyance corresponding to the levels was also predicted to investigate the subjective magnitude of the problem. NMTs are measurement stations located at and around the Vancouver airport, which record A-weighted, 1-s, equivalent-continuous noise levels. The GFPE model predicted levels in third-octave bands from 25 to 4000 Hz. The total A-weighted GFPE calculated levels, and the corresponding measured A-weighted, equivalent-continuous NMT levels during the run-ups, are plotted in Fig. 8.

Source-radiation characteristics were only available for the three aircraft discussed earlier; the Beechcraft-1900 and Dash-8 propeller aircraft, and the CRJ200 jet aircraft. Since, however, run-ups at YVR are not limited to these three aircraft, when predicting run-ups that involved other aircraft, the data for one of the three measured aircraft were "substituted." The aircraft noise "assigned" to it was chosen based on similarity of the aircraft type (jet/propeller) and engines (number/type). The substitutes (e.g., a Dash-8 for a D2, a CRJ200 for a Boeing 767, etc.) were not necessarily good matches. Clearly, the cases for which the source sound levels were unknown were associated with an increased prediction uncertainty. Moreover, the orientation of the aircraft (for directivity) was not always recorded in the run-up incident report. Finally, the actual power setting of the aircraft was not always as indicated on the incident report if, for instance, the operator decided to test at other power settings (see the examples in Secs. IV C 1 and IV C 4). As was shown in Sec. III B, these uncertainties can lead to large variations in predicted levels. Thus, in cases for which information was missing, GFPE predictions used estimated values of the missing

inputs. One or two sets of plausible values were used, to predict levels indicative of the ranges predicted; these are shown in Fig. 8. The cases labeled "low" are either for the single level, or the lowest of two values, predicted; the cases labeled "high" are for the highest of two values predicted. The prediction that was closest to the average level was used as the basis of the discussion.

One of the run-up locations at Vancouver International Airport had a blast fence located at the southern end. Previous work had measured the insertion loss of this particular blast fence.[5] While the insertion loss varied somewhat across the frequency spectrum, for the midfrequencies of most interest here it was relatively flat, with an average value of approximately 8 dB. Thus, for cases in which the blast fence was involved, an attenuation of 8 dB was subtracted from the predicted levels.

1. Available aircraft source levels, known headings

There were three events on two nights for which atmospheric conditions were stable, the correct aircraft source levels were available, the aircraft headings were known, and the wind speeds were below 5 m/s (in fact, they were all below 4 m/s): one on 19 February 2005, and the two on 18 May 2006 (A and B). In the first two cases, GFPE predictions overestimated the measured levels by 9 and 5 dBA, respectively. However, for the run-up on 19 February 2005, the blast fence was between the aircraft and the receiver. Applying the average insertion loss, the predicted value for 19 February 2005 is within approximately 1 dB of measurement.

8 September 2005 also had a known heading and available aircraft-noise data, but took place during neutral conditions. This run-up was recorded as at "idle" engine power. In

the past, there have been problems with requests for run-ups at a lower power setting, when in fact the aircraft were run at a higher power setting. Following a suspicion that a higher power setting may have been used for this runup, GFPE predictions were made for both idle and full power. Full-power predicted levels were 7 dBA below the average measured level, compared to the idle-power level, which was 23 dBA below the average measured level. Both predicted levels are shown in Fig. 8; however, due to the likelihood that the aircraft was run at full power, the full-power level was used for further analysis.

2. Unavailable aircraft source levels, known headings

Considering now cases for which the source data were unavailable for the specific aircraft but the heading was known, two stable-atmosphere cases (with the same aircraft—the events took place one after the other during the same night) and one neutral-atmosphere case were available with wind speeds below 5 m/s: on 25 February 2006 (A and B) and on 4 March 2005, respectively. For the stable cases, the GFPE predictions were 12 and 16 dBA below the average levels. The cause of these disagreements is likely the aircraft substitution of the Beechcraft 1900 for the Beechcraft 100.

In the neutral case, on 4 March 2005, levels were over-predicted by 17 dBA relative to the average measured level. Here again, however, the blast fence was located between the source and receiver. When including the attenuation due to the blast fence, the GFPE prediction is 9 dBA above the measured values.

3. Available aircraft source levels, unknown headings

The need to estimate the aircraft heading (orientation) when it was not recorded makes accurate prediction difficult, since the data available suggest that aircraft, particularly propeller aircraft, can be directional. As discussed in Sec. II, there is considerable uncertainty associated with the noise levels behind aircraft—particularly jet aircraft—adding to the complexity of predictions in cases missing relevant information. If no information was available, a source level at the front of the aircraft, and a source level toward one of the rear sides of the aircraft were used. For the run-ups on one night (5 March 2006 A and B), levels were predicted at two NMTs for aircraft for which source levels were available, but the headings were unknown. These predictions were 20 and 14 dBA above measured levels at Position 0 (see Fig. 1—levels at Position 5 were overpredicted by even more). Taking into account the combination of the insertion loss of the blast fence, and the possibility that the true heading resulted in lower sound-pressure levels, this prediction may in fact be much closer to the measured level. 5 March 2006 was also noted to have a stable atmosphere. The positive change in temperature between heights of 1.6 and 6 m was the largest that occurred for all of the run-up events considered: 4.3 °C over 4.4 m. Wind speeds were relatively low, at 1.1 m/s at 10 m height; the difference between prediction and measurement is not believed to be caused by unknown atmospheric-parameter values in this case.

4. Unavailable aircraft source levels, unknown headings

Four run-up events had unknown aircraft source levels and unknown headings: two in stable conditions, 17 July 2005A and 25 February 2006C, and two in neutral conditions, 27 July 2005 and 23 April 2006. Surprisingly, GFPE predictions underpredicted the average levels in the four cases by only about 1, 2, 1, and 3 dBA, respectively. The event on 17 July 2005 was recorded as an idle event, but when the full-power source data were used, the best prediction accuracy was obtained. The idle power setting listed for 27 July 2005 gave good agreement—better than the full-power setting. This suggests that sometimes the aircraft being run-up was at the idle-power setting as recorded, but that at other times it was at a higher power setting. The 23 April 2006 prediction gave a level equal to the background-noise level. What is interesting here is that the decibel sum of the background-noise and predicted levels equals the measured level. The prediction for 25 February 2006C is well above the background levels, below the peak level, and only 2 dBA different from the measured level. It is surprising how well the levels in this section agree, given the uncertainties in the inputs.

5. Wind speeds above 5 m/s

Four run-ups, each of which occurred during high-wind conditions, remain to be discussed: these occurred on 17 July 2005B, 2 May 2006, 12 October 2005, and 6 November 2005. If wind speeds had been lower, the former two run-ups would have been categorized as stable, and the latter two as neutral. If those conditions are assumed, GFPE predictions are about +16, +6, −1, and +1 dBA different from the measured levels of the noise events. Wind should not affect the propagation of sound significantly in any of these cases, because the source-receiver direction was not parallel to the wind direction (the differences between the source-receiver directions and the wind directions in the four cases were around 55°, 124°, 74°, and 135°, respectively). These stronger winds are, however, expected to affect the atmosphere, preventing the "standard" temperature profiles (parabolic or linear) from existing at greater heights, or even from being established. What makes this analysis extremely difficult is that information pertaining to the aircraft source levels, and/or the aircraft heading, is missing for all four run-up events, except on 6 November 2005 (for which the measured level was overpredicted by only 1 dBA). While, in principle, strong winds should affect the atmosphere in such a way as to make predictions difficult, uncertainties associated with the aircraft source data make this difficult to isolate. The fact that more than one GFPE prediction (i.e., the high and low predictions in Fig. 8) was made for all events except 17 July 2005B increases the chance of finding one prediction result that agrees well with the measured level.

6. Summary

The prediction results, corrected for blast-fence insertion loss where applicable, have been replotted against the measured levels in Fig. 9. The many points lying near the $y=x$
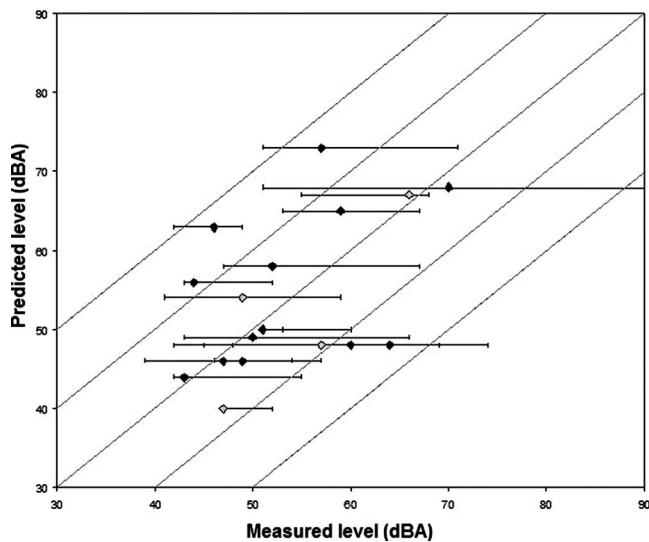
FIG. 9. Comparison of total, A-weighted GFPE predictions with NMT measurements. The central diagonal line represents $y=x$, the other lines are at 10 and 20 dBA offsets: ($\diamond$) Predictions for which all information was known. ($\blacklozenge$) Predictions for which some information was unknown. Error bars indicate the range of measured NMT levels, from background to peak.

line suggest that the GFPE model generally predicts measured sound-pressure levels with reasonable accuracy. The fact that this sometimes occurs when the input data are not well known apparently indicates that good agreement can sometimes be a coincidence. The symmetrical distribution of the data indicates that the GFPE model does not have an inherent tendency to over-or underestimate sound levels.

While many (71%) of the event-average levels in Fig. 9—after correction for barrier insertion loss, where applicable—were predicted within 10 dBA, some events were inaccurately predicted by nearly 20 dBA; the worst prediction, which was 17 dBA higher than the average NMT level, was for a Boeing 767 on 4 March 2005.

Given this small sample size (17), and the surprisingly accurate predictions discussed in Secs. IV C 4 and IV C 5, it is hard to draw conclusions regarding what conditions are more likely to result in accurate or inaccurate predictions. It appears that, for run-up events for which the aircraft noise levels are available, it is nearly three times (58%) more likely that prediction will be within 10 dBA than if the levels were estimated with a substitute aircraft (20%). It is more difficult to draw conclusions for the cases with known/ unknown headings, with wind speeds greater/less than 5 m/s, and with stable/neutral atmospheric states, as they all gave similar results. It is, of course, of interest to consider whether prediction had a tendency to be above or below the measured levels; in fact, overestimates appear to be equally as likely to occur as underestimates for predictions within 10 dBA of measured levels (42%), and for predictions within 20 dBA of measured levels (60%). Again, it is difficult to have high confidence in these overall results with this small sample size, and given the surprising prediction results of Secs. IV C 4 and IV C 5.

Though the data pool is limited, the results in Sec. IV C 4 suggest that if all information is available— particularly the noise levels of the aircraft performing the run-up—prediction can be accurate within about 10 dBA. With relevant information unknown, prediction uncertainty will increase; with reasonable choices, however, predictions can be expected to be accurate within about 20 dBA.

D. SEL and annoyance

It is of interest to investigate the annoyance associated with aircraft run-ups at Vancouver International Airport. Since night-time run-ups were of concern, annoyance was quantified in terms of the percentage of people awoken, as described in ANSI S12.9-2000/Part 6.[14] This standard relates the indoor SEL to the percentage of people awoken, as follows:

$$\% \text{Awakenings} = -7.02 + 0.14 \text{ SEL}. \tag{2}$$

Using the outdoor SEL predictions, the percentage of people annoyed (awoken) due to the engine run-ups identified in the present work ranges from 1.8% to 9.5%, with the average being 5.9%. An uncertainty of ±10 dB in SEL corresponds to an uncertainty of ±1.4%, ±20 dB to ±2.8% uncertainly.

Neither the indoor SEL nor the actual percentage of people awoken was known (no measurements were taken indoors). The only information that was available was that, at the time of the events in question, people were sufficiently annoyed by the noise that they contacted the airport with their concerns.

Indoor levels are equal to outdoor levels reduced by the insertion loss of the residence. Since these are difficult to estimate accurately, no attempt has been made to do so here. If indoor levels were available and used, as they should be according to ANSI S12.9-2000/Part 6, SEL would be lower; thus, the true percentage of people awoken is likely lower than as predicted earlier.

While light sleepers may be very sensitive to noise and be awoken regardless of the SEL, to lower this percentage theoretically to zero, SEL should be below 50 dB. Exceptions will always exist. Depending on the geometry of neighborhoods and of houses, some sound focusing may take place, making some areas louder than others. Even if predictions of SEL are well below 50 dB, there is a chance that some people will be awoken.

V. CONCLUSION

The objective of this work was to consolidate, evaluate, and then apply a prediction tool that could assist the Vancouver Airport Authority in managing aircraft-engine run-ups, to minimize noise levels and disturbance for the residents of the communities living near the airport. This objective has been achieved.

Parameters that influence the propagation of sound were consolidated into a simple outdoor sound-propagation model. The original GFPE model that was available when this work began was modified to better describe the ground and the atmosphere, and to produce a more useful output. To ensure that changes made to the GFPE code were accurate, the model was evaluated. A comparison of predictions with results in the literature found very good agreement—typically

within 2 to 3 dB. The agreement was considered sufficiently accurate to apply the GFPE model in realistic cases.

Availability of accurate input information was a big issue: a sensitivity analysis found that the state of the atmosphere, wind speeds, aircraft headings, and engine-power levels were the major sources of variability in predicted noise levels. Predictions were compared with noise levels measured at noise monitoring terminals near the airport. Some input parameters were difficult to obtain with accuracy, and estimates of unknown input data were required in several cases. For the four cases for which all input information was available, predicted levels were within 10 dBA. For cases where some information (i.e., the aircraft heading) was missing, the prediction error was within 20 dBA.

The sound-pressure levels predicted in the community for a sample set of 17 run-up events correspond to a predicted percentage of people awoken ranging from 1.8% to 9.5%. Identifying annoyance by percentage of people awoken is likely a better way to assess the number of people disturbed by noise than are complaints, as many members of the community do not feel inclined to report incidents, even if their sleep is disturbed.

While uncertainties of 10 dBA (and 20 dBA) are rather large, given the complexity of the problem and the small data pool, the agreement was good. Almost all prediction parameters—the ground impedance, the temperature and wind profiles, the source levels, the source and receiver positions, and the air absorption—had to be estimated or approximated in some way. The results of this study show that obtaining more accurate information—particularly relating to the aircraft noise-radiation characteristics and the atmosphere—is crucial to obtaining reliable predictions.

ACKNOWLEDGMENTS

[1] K. Attenborough, "Sound propagation close to the ground," Annu. Rev. Fluid Mech. **34**, 51–82 (2002).

[2] T. F. W. Embleton, "Tutorial on sound propagation outdoors," J. Acoust. Soc. Am. **100**, 31–48 (1996).

[3] A. Nakashima and M. Hodgson, "Effect of realistic grounds and atmospheric conditions on single-channel active control of outdoor sound propagation," J. Acoust. Soc. Am. **117**, 1080–1087 (2005).

[4] K. A. Scherebnyj, "Prediction of noise levels and annoyance from aircraft run-ups at the Vancouver International Airport," M.A.Sc. dissertation, Department of Mechanical Engineering, University of British Columbia, 2006.

[5] P. Germain, J. Guo, and M. R. Hodgson, "Measurement of propeller-aircraft run-up noise," Can. Acoust. **29**, 21–27 (2001).

[6] E. M. Salomons, *Computational Atmospheric Acoustics* (Kluwer Academic, Dordrecht, The Netherlands, 2001).

[7] E. M. Salomons, "Improved Green's function parabolic equation method for atmospheric sound propagation," J. Acoust. Soc. Am. **104**, 100–111 (1998).

[8] K. E. Gilbert and X. Di, "A fast Green's function method for one-way sound propagation in the atmosphere," J. Acoust. Soc. Am. **94**, 2343–2352 (1993).

[9] K. Attenborough, "Ground parameter information for propagation modeling," J. Acoust. Soc. Am. **92**, 418–427 (1992).

[10] K. Attenborough, S. Taherzadeh, H. E. Bass, X. Di, R. Raspet, G. R. Becker, A. Gudesen, A. Chrestman, G. A. Daigle, A. L'Espérance, Y. Gabillet, K. E. Gilbert, Y. L. Li, M. J. White, P. Naz, J. M. Noble, and H. A. J. M. van Hoof, "Benchmark cases for outdoor sound propagation models," J. Acoust. Soc. Am. **97**, 173–191 (1995).

[11] B. Gauvreau, M. Bérengier, P. Blanc-Benon, and C. Depollier, "Traffic noise prediction with the parabolic equation method: Validation of a split-step Padé approach in complex environments," J. Acoust. Soc. Am. **112**, 2680–2687 (2002).

[12] G. A. Daigle, J. Nicolas, and J.-L. Berry, "Propagation of noise above ground having an impedance discontinuity," J. Acoust. Soc. Am. **77**, 27–138 (1985).

[13] K. E. Gilbert, R. Raspet, and X. Di, "Calculation of turbulence effects in an upward-refracting atmosphere," J. Acoust. Soc. Am. **87**, 2428–2437 (1990).

[14] ANSI S12.9-2000/Part 6-American National Standard Quantities, "Procedures for description and measurement of environmental sound. Part 6. Methods for estimation of awakenings associated with aircraft noise events heard in homes" (Acoustical Society of America, Melville, NY, 2005).

Long-range acoustic scattering from a shallow-water mud-volcano cluster

Charles W. Holland[a] and John R. Preston
Applied Research Lab, The Pennsylvania State University, State College, Pennsylvania 16804

Douglas A. Abraham
CausaSci LLC, Arlington, Virginia 22205

Analysis of reverberation measurements in the Straits of Sicily shows high intensity, discrete, scattered returns 10–20 dB above background reverberation. These returns are due to scattering from mud volcanoes. The reverberation from the mud volcanoes at ranges of 15–22 km is reasonably consistent over these spatial scales (i.e., kilometers) and temporal scales of several hours; measurements separated by 4 years are also similar. Statistical characterization indicates that the reverberation associated with a mud-volcano cluster is strongly non-Rayleigh and that the reverberation can be characterized by a single (shape) parameter, roughly independent of frequency. The non-Rayleigh statistics, with a concomitant increase in the probability of false alarm, indicate that mud volcanoes are a likely source of clutter. Mean target strengths were estimated at 1–11 dB over 160–1400 Hz and are consistent with target strengths measured during a different year at short (direct-path) ranges. Accumulated evidence points to small (order 10 m diameter and several meters high) carbonate chimneys on the mud-volcano edifice as the scattering mechanism as opposed to the edifice itself or scattering from gas bubbles in the water column. Thus, the results represent acoustic scattering from mud volcanoes in a quiescent state.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2773995]

I. INTRODUCTION

The performance of active sonar systems in littoral environments is often limited by discrete clutter and diffuse bottom reverberation. Discrete clutter, which leads to high false alarm rates, is widely viewed as one of the most important limiting environmental factors in the operation of active sonar systems in littoral areas. Discrete clutter tends to be nearly ubiquitous, but its characteristics have been difficult to predict. One of the primary reasons that prediction has been difficult is that the underlying mechanisms giving rise to the clutter are largely unknown. In a very general sense, it is certainly understood that high scattering arises from anthropogenic features such as wrecks and well-heads (e.g., Refs. 1 and 2), biologics (e.g., Ref. 3), as well as geologic features (e.g., scarps and cliffs[4]). However, especially for the latter, there are many geologic features whose scattering mechanisms and scattering characteristics are not understood.

Strictly speaking, clutter can be defined as "target-like" returns that persist over a target-like track. Here, we define clutter as reverberation with strongly non-Rayleigh statistics or alternatively as high reverberation with relatively large target strengths (>0 dB) and temporal characteristics such that the time spread is not much larger than that induced by the channel.

In this study, the objective is to examine the statistical characteristics of long-range reverberation from a class of geologic features called mud volcanoes. Mud volcanoes form due to the rise of fluidized sediments and/or gas along a fault or on top of a sea-floor-piercing shale diapir. They may occur in sedimentary areas with hydrocarbon generation at depth, originate from thick clay beds, usually erupt along fault lines, and often bubble gas (mostly methane), and sometimes oil. They are known to occur in a variety of geologic settings including the abyssal parts of inland seas, active margins, continental slopes of passive margins, and continental shelves.[5] While deep-water mud volcanoes (e.g., Ref. 6) have been studied for some time, relatively little is known about the distribution and characteristics of shallow-water mud volcanoes. One reason for this is that shallow-water mud volcanoes tend to be much smaller scale than their deep water counterparts, of order $10^{0–1}$ m in height as opposed to 10^2 m for those in deep water, and thus less easily detected. Mud volcanoes sometimes occur in clusters: in this area, of order 10 MV within a 1 km^2 region.

In a previous study,[7] scattering from a mud volcano was measured at short range (a few 100 m), i.e., in the direct-path region. Target strengths at low grazing angles 3°–5° from coherent pings were measured at 4–14 dB over 800–3600 Hz. Low grazing angles were measured in order to be relevant to the long-range clutter problem, examined here. The evidence from that study, in terms of scattering mechanism, suggested that the scattering is associated with carbonate chimneys on the mud volcano as opposed to scattering from gas bubbles entrained in the sediment or emitted in the water column, or the mud-volcano edifice itself.

a)Electronic mail: holland.cw@psu.edu
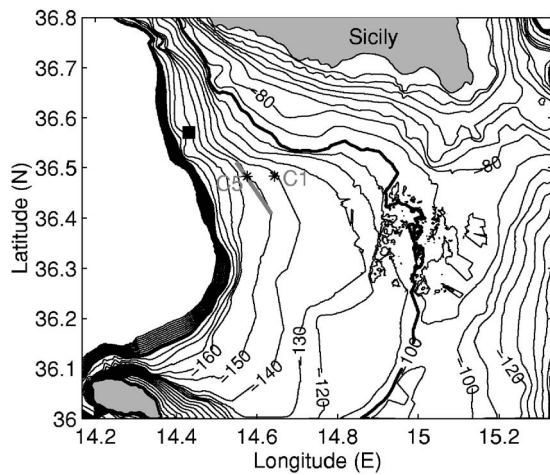
FIG. 1. Experiment area in the Straits of Sicily showing the location of the MV cluster (■) and source-receiver location of the two quasimonostatic reverberation experiments (∗), C1 and C5. The location of the transmission loss track is shown as a thick gray line crossing C5.

In this study, long-range (order 10 km) reverberation measurements from the same mud volcano, and from others within the same cluster, are presented in order to examine: (a) the statistics of the long-range reverberation, (b) the variability of the reverberation associated with time and spatial location of source-receiver, and (c) the frequency dependence of the target strength. The two measurement approaches, short range and long range, have distinct and somewhat mutually exclusive advantages. For example, measurements in the direct-path region have the advantage of reducing or eliminating uncertainties due to oceanographic variability and potentially poorly known geoacoustic properties between source and feature. Another advantage of the short-range measurements is the ability to directly probe the scattering mechanism, by determining the location of the scattering (especially in the vertical plane). However, the local measurements have more difficulty controlling aspect,[8] are difficult to perform at low frequencies (below about 400 Hz) and in order to study the statistics of the clutter (including all of the multipath effects), a long-range observation is required. In addition, the long-range measurements have many orders of magnitude greater spatial coverage and are easier to conduct. Ultimately, we see a promising strategy in using a combination of both short- and long-range observations.

In the following sections, the experiment area is first described along with an overview of the mud-volcano characteristics. Then, the reverberation observations are shown and discussed as a function of space and time. A statistical analysis of reverberation from the mud-volcano cluster is analyzed to characterizes the probability density function (PDF) of the reverberation envelope and a simple model is proposed that captures the frequency dependence of the statistics. The observations are analyzed in terms of target strengths, and the final section contains discussion and conclusions.

II. EXPERIMENT AREA DESCRIPTION

The Malta Platform (see Fig. 1) is a structural unit of the
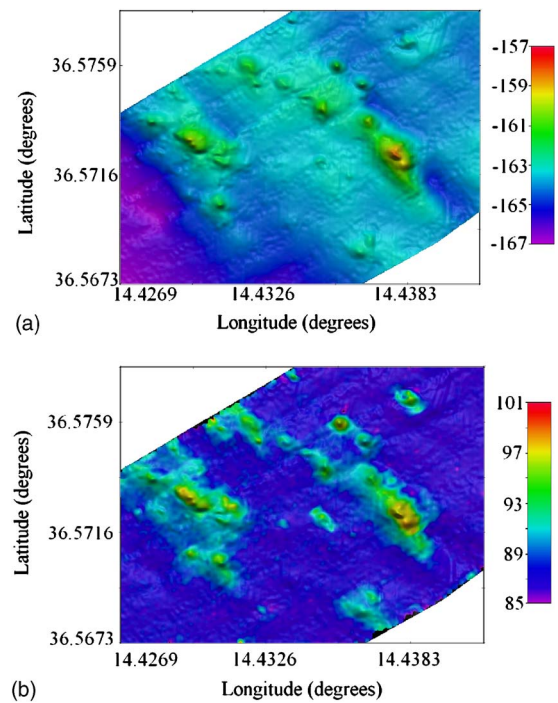
FIG. 2. (Color online) (a) Multibeam bathymetry showing location and height field of the mud volcanoes (in meters). The area is approximately 1350 × 1200 m. (b) A mosaic of values proportional to backscattering strength in decibels (arbitrary units), corrected to 70° (from Ref. 7).

African foreland, colliding with the European plate.[9] The mud volcanoes (MVs) of interest are located on the western margin of the Malta Platform and within the Scicli fault zone, where the main faults trend SW-NE. This area and the MV characteristics are described in more detail in Refs. 7 and 10.

By way of background, we present 240 kHz multibeam measurements of bathymetry and backscattering at the mud-volcano cluster [Figs. 2(a) and 2(b)]. Within the survey region, the average sea floor is gently sloped, from a depth of 164 m in the northeast corner of the survey to 167 m in the southwest corner. Several MVs are visible, rising to a maximum of approximately 5 m above the mean sea-floor depth. The outer length scales observed in these features are quite variable, ranging from 40 to 400 m. Also visible in the data are nearly linear depressions adjacent to some of the larger MVs that may be associated with faults and/or regions of fluid escape. The multibeam sounding resolution varied across the swath from approximately 4 to 10 m in the across-track direction, and 2.5 m in the along-track direction. In postprocessing, the data have been low pass filtered, so that the resolution of the data in Fig. 2(a) is approximately 10 m. This resolution does not resolve the ∼10-m-diameter carbonate chimneys (see Fig. 4 of Ref. 7) that are believed to be the dominant scattering mechanism.

In Fig. 2(b), the inherent (vertical) angular dependence in the backscatter has been removed (see Ref. 7) and the image represents backscatter that would be expected at a fixed grazing angle (70°). The backscattering data show three distinct regions: the background (∼87 dB) intermediate regions that are 2−4 dB above the background, and high backscatter regions, which are 6−10 dB above the background.

Holland *et al.*: Long-range scattering from mud volcano cluster   1947
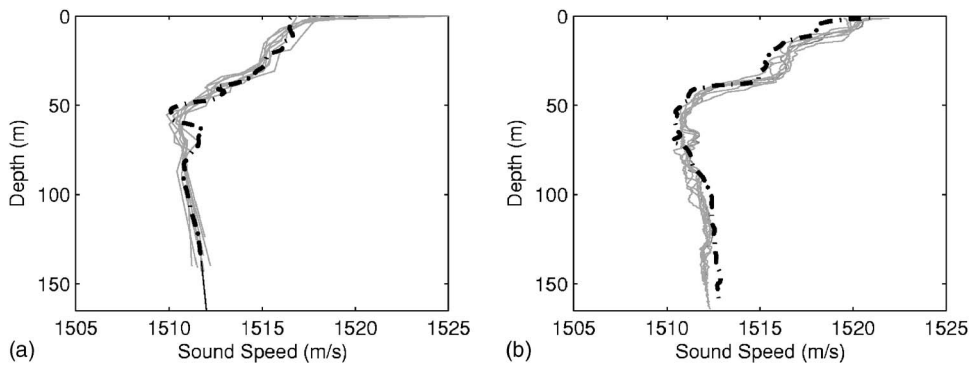
FIG. 3. Sound speed profiles during the reverberation experiments, XBT(—), CTD(—): (a) Site C1 to MV cluster, linear extrapolation of CTD to 165 m (---), the CTD was taken several hours before the reverberation experiment. (b) Site C5 to MV cluster, CTD was taken the day after the experiment.

These three regions appear to correspond to three classes of surficial sediments in the survey area: fine-grained sediments, coarse-grained sediments, and carbonate blocks and fragments, respectively.[7] The last category forms a useful, if inexact, metric for determining the mud-volcano locations. That is to say, that the location of the high scattering in Fig. 2(b) is a clearer representation of the position of the MVs than just the height field [Fig. 2(a)] by itself.

In a later section (Sec. V) a target strength analysis is performed on reverberation returns from the largest mud volcano in the cluster (at 36.5716N 14.4383E). This MV is roughly elliptical in shape with dimension 150×50 m and oriented at 145° (re North). Direct-path target strength measurements were previously conducted[7] from this same feature.

III. REVERBERATION OBSERVATIONS

Reverberation from the mud volcanoes is presented from two different source-receiver positions (see Fig. 1) collected in two different years: Site C1 in 2000 and Site C5 in 2004, which are ~21 and ~15 km from the MV cluster, respectively. Both sets of measurements show consistent reverberation returns over time scales of hours as well as being consistent with each other. This is indicative of a scattering mechanism that is stable over those time scales.

A. Source-receiver at Site C1

Long-range reverberation data from Site C1 to the MVs (see Fig. 1) were acquired 3 May 2000 during the Boundary 2000 experiment.[11] The equipment employed and the ocean environment are briefly described in the following. More details can be found in Ref. 12.

Impulsive sources, Mk82 Sound Undersea Signals (SUS), were launched from the vessel and initiated within a few hundred meters of the receiver at 91 m depth. Three SUS were launched on five legs, where each leg was at a different bearing and the SUS were dropped within a few kilometers of Site C1. The receiver was a three-aperture nested horizontal array of 128 elements per aperture towed at a depth of ~75 m. The data were digitized at 6000 Hz sampling rate and low-pass filtered at 1780 Hz with a seven-pole six-zero elliptic (70 dB per octave roll-off) antialias filter. Data presented here were time-domain beamformed (on the aperture with 0.5 m spacing) with Hann shading and 129 beams, so that at the design frequency, 1500 Hz, the beams

overlap at the 1.5 dB down points. In the processing, an integration time of $\tau = 1/12$ s was employed, which corresponds to a radial patch dimension of 60 m. The data were incoherently averaged (amplitude squared) over frequency in selected 100 Hz bands from 100 to 1800 Hz. The geometry is such that the azimuthal scattering angle is oriented ~35° relative to the axis of the main MV.

During the reverberation experiment, the wind speed varied between 5 and 10 knots and the rms sea surface wavelength measured on a waverider buoy varied from 0.15 to 0.25 m. Sound speed profiles were collected using expendable bathythermographs (XBTs) every ~30 min during the experiment from the ship as it executed the reverberation pattern. The sound speed profiles [see Fig. 3(a)] exhibit a very minor variability in space and time. The profiles were extended in depth by ~10 m by linear extrapolation to accommodate the slightly deeper depths at the MV cluster. It is somewhat reassuring that the gradient thus used in the extrapolation is very similar to the slight positive gradient measured at the MV cluster at the same season but different year (2004).

The range from source to scatterer was approximately 21 km and the bathymetry is gently sloping to the west (downslope from source to the MVs). Given the short distance between source and receiver (order 100 m) relative to the distance to the MVs, the geometry is assumed to be monostatic in the analysis.

Observations of the clutter along different bearings and at various positions permit us to resolve the left-right ambiguity inherent in the towed array data and also permit us to qualitatively examine the variability of the clutter in time and space. An example of reverberation beam time series data is shown in Fig. 4. The beam intersects with the MV cluster at 27.5 s and several distinct returns are observed at that time, about 10–15 dB above the background reverberation. At this frequency and bearing, the beamwidth is about 1.5° and the illuminated patch size at the MV cluster in the cross-range dimension is about 550 m. The reverberation from 100 Hz bands at other center frequencies is shown in Fig. 5, where significant returns associated with the MV cluster are observed at all frequencies.

Generally, it is observed that the gross arrival structure is roughly constant from ping to ping. This is seen, for example, from a ping initiated along the same heading but 1.5 km further north (Fig. 6) and about 13 min earlier in time. Sometimes, however, there is a close correspondence to
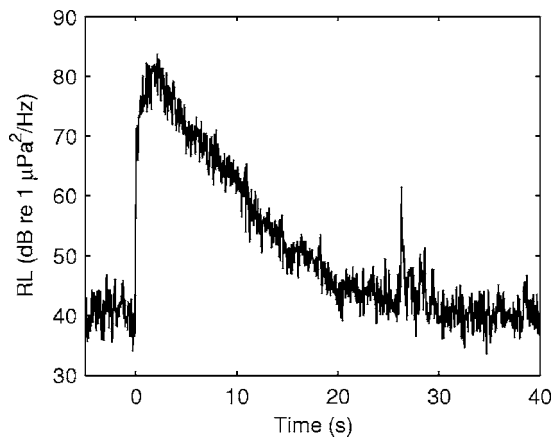
FIG. 4. Reverberation beam time series at 1400 Hz from Site C1. The reverberation from the largest MV is at 27.5 s, and is 10–15 dB above the background noise/reverberation. The vessel bearing was 180° (leg2) and the MV cluster is at a bearing of 297°. The receive array was linear, i.e., composed of single elements.

individual reverberation peaks in time and amplitude, when the data are time shifted relative to a prominent arrival. An example of this shown in Fig. 7, where Fig. 7(a) shows a comparison for different positions along the same bearing and Fig. 7(b) shows a comparison for different positions and bearings. Strong differences in Fig. 7(b) at times other than 27.5–28.5 s are most likely due to differences in the ambiguous beams.

B. Source-receiver at Site C5

Reverberation data from Site C5 (see Fig. 1) are presented mainly as a confirmation of the results at Site C1. These data were acquired 20 May, 2004 during the Bound-

ary2004 experiment. The experiment was similar to that conducted at C1, e.g., three SUS were launched on five legs, where each leg was at a different bearing. The main equipment difference between this experiment and that of 2000 was a cardioid towed array, which allows left-right discrimination (typical rejection of ~15 dB). Also, in this experiment a coherent source was employed.

The cardioid array had 84 triplet elements with interelement spacing of 0.42 m towed at a depth of ~90 m. Both impulsive (SUS) and coherent sources were employed. The SUS were deployed at 91 m and the coherent source at 90 m. Data presented here were time-domain beamformed with Hann shading. The processing was similar to that at Site C1, i.e., with an integration time of $\tau=1/12$ s, incoherently averaged in 100 Hz bands. The sources were all deployed in the vicinity of Site C5, such that the azimuthal scattering angle is oriented roughly 28° relative to the axis of the main MV.

During the reverberation measurements, the wind speed varied from 3 to 12 kn and the rms waveheight from 0.10 to 0.25 m. Sound speed profiles were collected using XBTs every ~30 min during the experiment from the ship as it executed the reverberation pattern [see Fig. 3(b)] and are similar to those at Site C1 in 2000, exhibiting only a modest variability in space and time.

Data from one of the SUS events are shown in Fig. 8. Note the peak levels across this frequency band are all roughly 15 dB above the background reverberation/noise and that the levels at 1000 and 1400 Hz are comparable to commensurate frequencies in the Site C1 data. The cause of the large peaks in the reverberation match-filtered time series at 12 and 15 s is unknown at this time, and may be due to other mud volcanoes or carbonate mounds. Match-filtered rever-
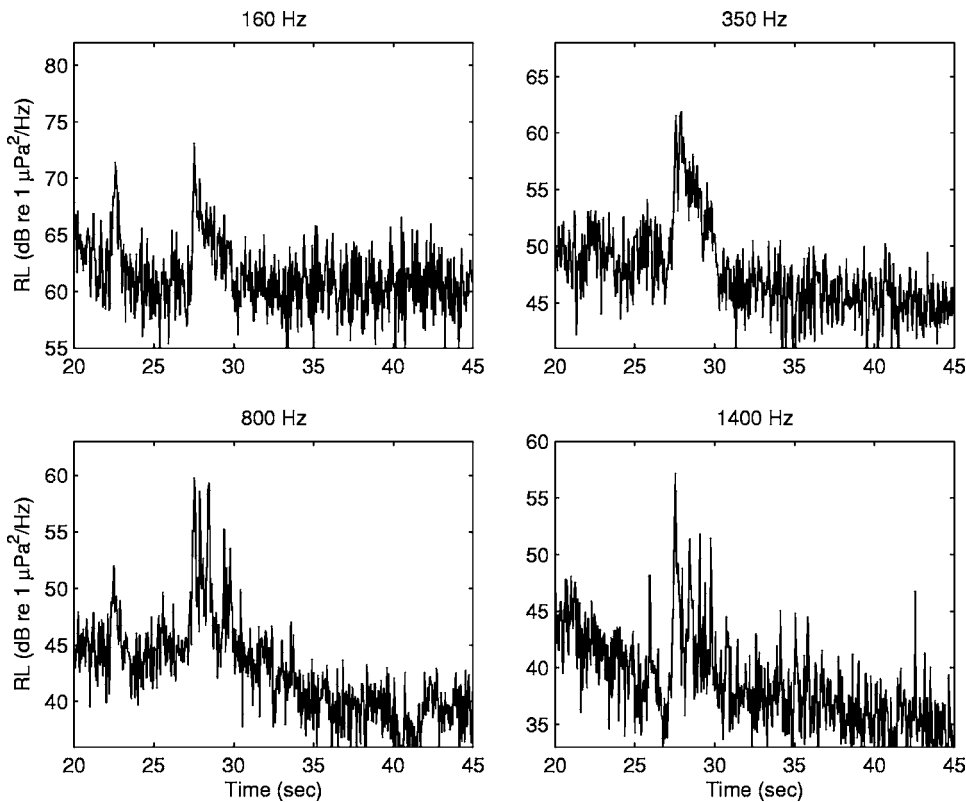
FIG. 5. Reverberation beam time series data from Site C1, leg 2 event 3. The reverberation from the largest MV is at 27.5 s, and is 10–15 dB above the background noise/reverberation. The features responsible for the returns at other times (e.g., 22 s) are not known at this time. The receive array was linear, i.e., composed of single elements.
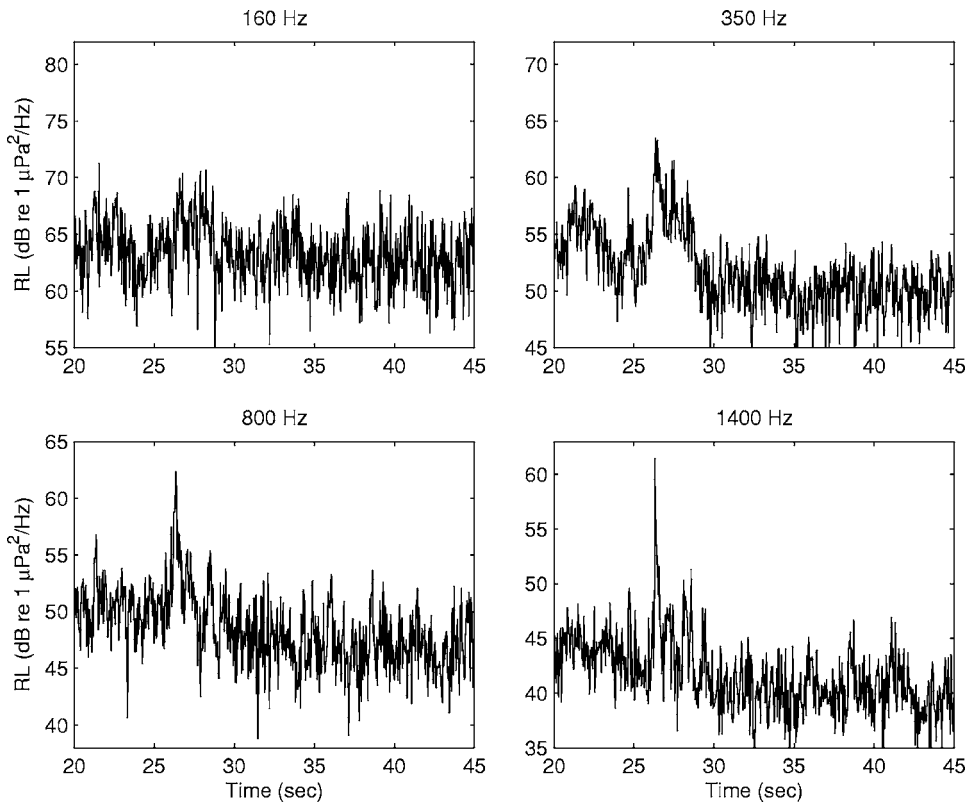
FIG. 6. Reverberation beam time series data from Site C1, leg 2 event 1 (1.5 km further north than that in Fig. 5). The reverberation from the largest MV is at 26.3 s. The receive array was linear, i.e., composed of single elements.

beration time series to the same MV using a coherent pulse (1.5 s 700–1700 Hz LFM) at a different range are shown in Fig. 9. The reverberation from the MV at 21.6 s is about 15 dB above the background.

In summary, the observations show consistent returns from the mud-volcano cluster. The characteristics of the returns are robust, i.e., not greatly sensitive to precise location, time, or waveform.

IV. STATISTICAL ANALYSIS

The following presents a statistical analysis of the mud-volcano echoes in 100 Hz bands at various center frequencies and proffers a statistical model to represent the echo envelope PDF. The 100 Hz bands are representative of what a coherent-source active sonar system might utilize. A statistical characterization of the envelope of the mud-volcano echoes could be used to improve modeling of active sonar performance in the presence of mud volcanoes and simulation of active sonar clutter representative of mud-volcano echoes.

A. Probability of false alarm analysis

From a statistical perspective, the reverberation time series described in Sec. III (e.g., Fig. 4) visually appear to be mostly Rayleigh with some spikiness representative of active sonar clutter, particularly in the region of the mud-volcano echoes. To quantify this, the probability of false alarm (P_{fa}) is estimated from 1 s snippets of data centered at the time of the mud-volcano echoes from four SUS shots[13] at Site C1 as well as from (arbitrarily) 2.5 s after and 5 s prior to each SUS shot. Before estimating P_{fa}, the data are filtered to 100 Hz bands with center frequencies of 160, 350, 800, and 1400 Hz (the center frequencies were chosen arbitrarily across roughly a decade of frequency). The envelope of the bandpass-filtered data is formed and then the time-varying

(a)   (b)
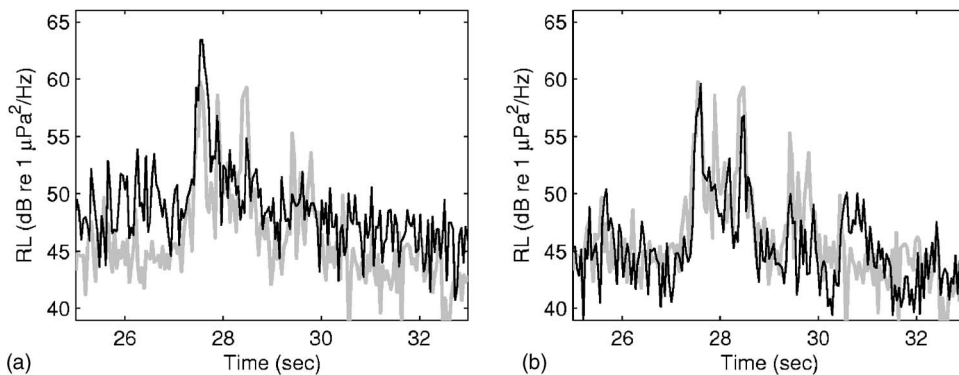
FIG. 7. Comparison of reverberation beam time series at 800 Hz from three shots at Site C1. The time base has been shifted to account for the slightly different ranges from the feature. Leg 2, event 3 is shown in the gray curves in both plots: (a) leg 2, event 2 (black line) and (b) leg 4 event 3 (black line). The two events in (a) are separated by 10 min, 1.2 km, and ship bearing of 1°; the two events in (b) are separated by 148 min, 0.03 km, and ship bearing of 73°. The receive array was linear, i.e., composed of single elements.

Holland *et al.*: Long-range scattering from mud volcano cluster
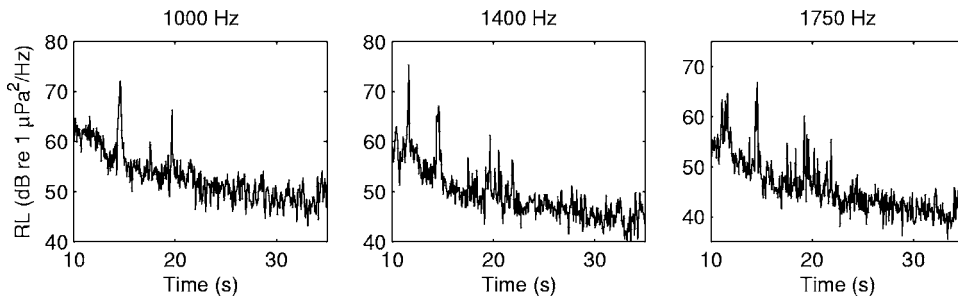
FIG. 8. Reverberation beam time series from Site C5 (see Fig. 1). The reverberation from the largest MV is at 19.7 s, and is ~15 dB above the background noise/reverberation. The vessel bearing was 179° (leg 4) and the MV cluster is at a bearing of 304°. The receive array was a cardioid, i.e., composed of triplet elements, which allows resolving of the left-right ambiguity.

background reverberation and ambient noise power level is removed using a sliding-window cell-averaging constant-false-alarm-rate normalizer[14] with 0.25 s auxiliary data windows placed 0.075 s before and after each test cell. The data from the four frequency bands and four SUS shots are then combined to estimate the P_{fa}. As seen in Fig. 10, the P_{fa} estimated from the mud-volcano echoes is extremely heavy tailed compared with the P_{fa} estimated from the earlier time window, which is closer to the Rayleigh probability density function (PDF), though still heavy-tailed. As is often the case, the ambient noise observed prior to the SUS charge is close to the Rayleigh PDF. From the P_{fa} curves shown in Fig. 10, it is clear that the mud-volcano echoes as well as echoes from other scatterers in the area will produce false alarms in sonar detection algorithms.

B. Statistical modeling of mud-volcano echoes

Other researchers have noted non-Rayleighness in reverberation, for example, Ref. 15 presented data from exposed basaltic relief near the mid-Atlantic Ridge. They concluded that, in that environment, scattering from randomly distributed facets of the same scale as the wavelength gave rise to the non-Rayleigh reverberation. Reference 1 identified rock outcrops in the Malta Plateau as a source of active sonar clutter and hypothesized natural sediment disturbances and changes in sediment type as other possible geologic sources.

To develop an appropriate statistical model for scattering from the mud volcanoes, we first consider the multibeam bathymetry data found in Fig. 2(a). The data are interpolated to a 5 m resolution grid and used to estimate the distribution

function of mud-volcano height above the sea floor (i.e., the height field or rough surface) in terms of an exceedance probability. This is presented in Fig. 11 as the probability of observing a depth shallower than that noted on the abscissa. For depths shallower than 164 m, the exceedance distribution function (EDF) is seen to closely follow an exponential distribution. An exponential distribution of heights is not unusual in nature, for example, see Refs. 16 and 17 for arctic sea-ice keels. An exponentially distributed sea-floor height was used in Refs. 18 and 19 to evaluate the statistics of reverberation from a rough surface in a waveguide through simulation and theoretical analysis. The reverberation from the exponentially distributed sea-floor heights was found[18] to be well represented by the K distribution.[20]

However, in the case of mud volcanoes the scattering appears to be dominated by carbonate chimneys (see Sec. V C), which are of a smaller scale than probed by the multibeam bathymetry data. It is plausible that the heights of the carbonate chimneys are also exponentially distributed (self-similar statistics are not uncommon in geological surfaces). Based on the observation that the scattering appears to be roughly independent of frequency over a decade and independent of azimuth, a field of randomly sized spheres (representative of the squat carbonate chimneys) may be a more appropriate (though admittedly simplistic) model than the rough-surface model. Based on the results of Reference 21, such a field of discrete scatterers with an exponentially distributed size would give rise to a K-distributed reverberation
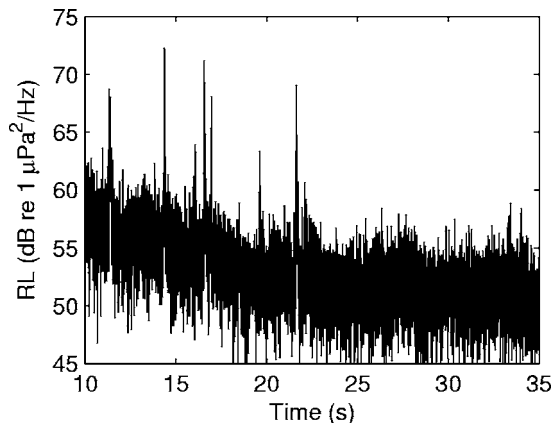
FIG. 9. Reverberation beam time series from Site C5 (see Fig. 1). The source pulse was a 1.5 s 700–1700 Hz LFM. The reverberation from the largest MV is expected at 21.6 s.
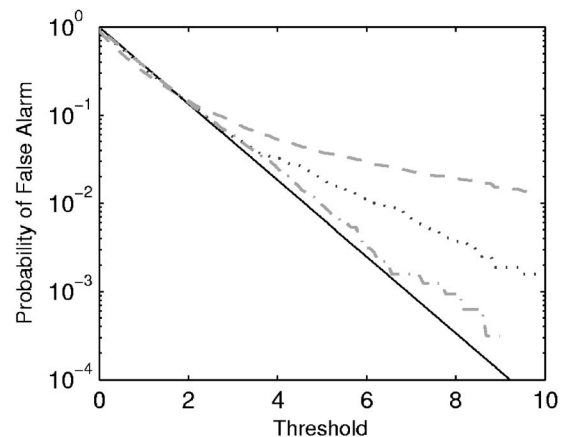
FIG. 10. The probability of false alarm is shown for, in order of increasingly heavy PDF tails, the Rayleigh PDF (solid black line), ambient noise (dash-dot line), reverberation received from 2 to 3 s after the SUS charge (dotted line), and mud volcano echoes (dashed line). The threshold is shown in linear scale (not decibels).
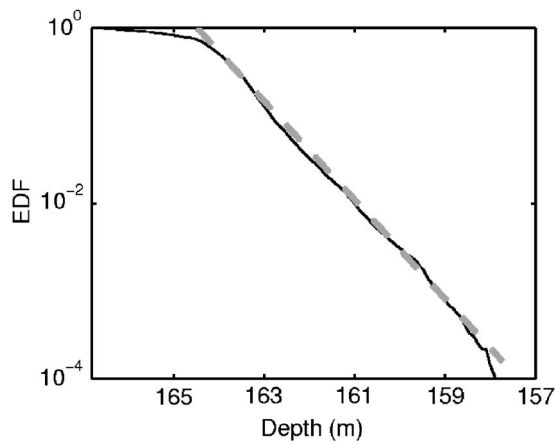
FIG. 11. Probability of observing a depth shallower than that noted on the abscissa [i.e., the exceedance distribution function (EDF) of the bathymetry] in mud volcano area, illustrating a very good fit to an exponential EDF (dashed line) shallower than approximately 164 m depth. The deeper depths are primarily associated with the northeast and southwest regions of the survey region where the bathymetry is gently sloping.

FIG. 12. K-distribution shape parameter estimates (tick marks) from mud volcano echoes are shown for four SUS shots before (black) and after (gray) accounting for the varying mud volcano echo energy to background diffuse reverberation and ambient noise power ratio. Data are only shown for cases where this ratio is greater than −10 dB with the following key x=Leg2Ev1, ○=Leg2ev2, □=leg2Ev3, ◇=leg4ev3. The solid and dashed lines represent the median over the displayed data.

envelope with the shape parameter proportional to the number of scatterers within the sonar resolution cell.

The fit of the K distribution to the mud-volcano reverberation is evaluated by analyzing the data obtained from the same four SUS events at Site C1. The array data are beamformed and the beam with the strongest mud-volcano echo is chosen from which 1 s windows of data centered about the mud-volcano echoes are taken for analysis. Prior to statistical analysis, the envelope of 100 Hz bands with center frequencies varying from 100 to 1400 Hz in 100 Hz increments is formed and the time-varying average background power level is removed using the previously described normalization process. A Kolmogorov-Smirnov (KS) test[22] was applied to the data to determine how well they fit the Rayleigh and K distributions. The PDF of the K distribution is

$$f_Y(y) = \frac{4}{\sqrt{\xi}\Gamma(\gamma)}\left(\frac{y}{\sqrt{\xi}}\right)^\gamma K_{\gamma-1}\left(\frac{2y}{\sqrt{\xi}}\right), \tag{1}$$

where y is the reverberation envelope, γ is the shape parameter, ξ is the scale parameter, and K is the modified Bessel function of the second kind. A Rayleigh PDF with the same average intensity is

$$f_y(y) = \frac{2y}{P}e^{-(y^2/P)}, \tag{2}$$

where the power is $P=\xi\gamma$. Holding the product $\xi\gamma$ constant, the K distribution simplifies to the Rayleigh distribution as $\gamma \to \infty$ indicating that large values of γ result in a distribution very similar to Rayleigh while small values of γ result in clutter-like, heavier-tailed distributions.

The asymptotic p value of the KS test statistic was used to accept or reject the data as being well fit by these two distributions at the 0.01-significance level (i.e., the probability of rejecting the fit when it is correct is 0.01). Of the data analyzed, 35 of 56 samples (63%) were fit well by the Rayleigh distribution and 54 of 56 (96%) were fit well by the K distribution. It should be noted that the power of this test (i.e., the probability of rejecting the fit when the PDF model

is incorrect) is limited owing to the small time-bandwidth product (=100) of each data sample. The data fit well by the Rayleigh distribution came from the lower center frequencies (500 Hz and below) where the mud-volcano echoes were less distinguishable from the background (diffuse reverberation and/or ambient noise).

As previously mentioned, the shape parameter of the K distribution is indicative of how heavy-tailed the reverberation envelope PDF is compared with the Rayleigh PDF. It also provides a link from the statistics of reverberation to physical mechanisms of scattering[21] and has recently been used to infer the spatial density of independent scatterers on the sea floor.[23] Based on the target strength estimates described in Sec. V, which are roughly constant over the frequency band being evaluated, the number of independent scatterers, and therefore the K-distribution shape parameter, is not expected to vary with center frequency. The K-distribution shape parameter is estimated from each of the aforementioned 1 s data windows using the bootstrap technique described in Ref. 24. The results are seen in the upper portion of Fig. 12 where a decreasing trend from a high of near 20 to a low near 0.2 is observed as center frequency increases, contradicting the expectation of a constant shape parameter. The solid line is the median of the shape-parameter estimates from the four SUS shots as a function of center frequency. The regions of high shape parameter correctly correspond to the regions where the KS test indicated a good fit to the Rayleigh distribution. The lowest values (occurring at the highest center frequencies) are representative of very heavy-tailed non-Rayleigh reverberation and would clearly cause false alarms in a sonar detection algorithm.

Based on the results of Ref. 21 the decreasing shape-parameter trend might be expected to arise from the decrease in the receive-array beamwidth causing a decrease in the number of scatterers contributing to the received signal at higher frequencies. However, the analysis of Ref. 21 assumes spatial isotropy in the scattering, which is not the case here.

Owing to the spatial compactness of the mud volcanoes relative to the sonar resolution cell cross range, decreasing the beamwidth (as arises from an increase in center frequency for a towed array) does not remove mud-volcano echoes in the received signal, but results in less diffuse reverberation and ambient noise and therefore a higher mud-volcano echo-to-background power ratio. As such, the data, prior to forming the envelope, may be more appropriately modeled by the sum of a K-distributed component representing the mud volcanoes and a complex-Gaussian component representing the background diffuse reverberation and ambient noise. The higher order moments of this combination may be used to form an equivalent shape parameter,[25]

$$\gamma_X = (1 + 1/\eta)^2 \gamma_0, \tag{3}$$

where γ_0 is the shape parameter of the K-distributed component and η is the power ratio of the K-distributed component (the mud-volcano echoes) to the Gaussian component (the diffuse reverberation and ambient noise). Choosing η based solely on the beamwidth of the array did not significantly reduce the spread of the shape parameter estimates observed in Fig. 12, perhaps owing to frequency-dependent ambient noise power levels. Therefore, in this preliminary analysis, η is estimated from the data in order to evaluate γ_0, which could then be used in performance modeling or simulating mud-volcano echoes. As a 1 s window of data is used to estimate the K-distribution shape parameter, η is approximated by the ratio of the energy in the data samples exceeding a threshold (those assumed to arise from the mud volcano) to the total energy in the data window when the mud-volcano samples (the threshold crossings) are replaced by the threshold value. Since the mud-volcano echoes were not visible at many of the lower frequencies (see the analysis presented in Sec. V C), the threshold was chosen by first finding the maximum normalized value for each center frequency and then choosing the smallest one as the threshold. When the resulting value of η was greater than -10 dB, it was used to compute γ_0 as found in the lower portion of Fig. 12 along with the median (dashed line) over the samples at each center frequency. The larger variation at the lower center frequencies arises from the lower values of η observed at these frequencies as well as the increased variance expected for estimates of higher shape-parameter values (for large values the Cramer-Rao lower bound for the shape parameter is proportional to the shape parameter squared[21]). At the higher center frequencies (600 Hz and above), the shape parameter estimates exhibit less spread with 27 of 31 (87%) lying between 0.1 and 0.2. Such consistency supports the expectation that the shape parameter does not vary over these center frequencies and that these mud-volcano echoes result in a K-distribution shape parameter on the order of 0.1 to 0.2 for 100 Hz bandwidth processing. This conclusion is also supported by analysis with varying window sizes where the estimates of γ_X and η change but the estimates of γ_0 predominantly remain in the 0.1 to 0.2 range. Therefore the model comprising both K- and Rayleigh-distributed components may be used to examine sonar performance or simulate data from mud volcanoes where the mud-volcano shape parameter is taken as constant and quite small.

C. Discussion

The analysis in this section has been restricted to 100 Hz bands owing to their relevance to coherent-source active sonar systems and results in estimates of γ_0 in the range 0.1 to 0.2. This may be interpreted as the average number of mud volcanoes contributing to the sonar resolution cell for the given array and processing bandwidth and includes any potential effect of propagation. The results of Ref. 21 indicate that, in the absence of any propagation effects, one scatterer with an exponentially distributed size contributing to the sonar resolution cell would result in a K-distribution shape parameter of 0.5. Estimates smaller than 0.5, as observed in the mud-volcano data, can arise from nonstationarity in the sampling window or discrete scatterers following a size distribution function heavier tailed than the assumed exponential PDF. Nonstationarity in the sampling window certainly does arise from anisotropy in the spatial distribution of mud volcanoes, evident in Fig. 2 where the gentle slope in the southwest and northeast regions results in a reduced backscattering strength. Coupled with the significant cross-range extents of the sonar resolution cell (approximately 500–4000 m for the data considered), measured reverberation is clearly not from a homogeneous spatial distribution of mud volcanoes.

However, neither is there firm evidence that the size of the carbonate chimneys follows an exponential PDF. Despite the exponentially distributed mud-volcano heights, the carbonate chimney sizes may in fact follow some other super-exponential distribution (i.e., a distribution with heavier tails). It is also possible that the heavy tails in the PDF arise from over-resolution of a scatterer by the sonar resolution cell when the scatterer is not significantly larger than a wavelength, which occurs in the down-range dimension where the 100 Hz-bandwidth processing results in a 7.5 m cell size, the carbonate chimneys are approximately 10 m in diameter, and the wavelengths range from 1 to 9 m. Thus, in the current analysis, both explanations for the small shape parameter values are tenable and currently not resolvable without additional information.

Although propagation in a waveguide is expected to increase the K-distribution shape parameter as compared with direct-path scattering (see Refs. 18 and 26), the spatial compactness of the mud-volcano field should result in a minimal difference between a direct-path and long range estimate. Essentially, the increase in the shape parameter arises from multiple, statistically independent parts of the bottom combining coherently as a result of multipath propagation. The spatial compactness of the mud volcanoes is expected to dominate contributions from other parts of the sea floor, which essentially act to reduce η, but not affect γ_0 in the proposed model. This is supported by the rapid decay of the multipath echoes (discounting the SUS bubble pulse) seen in Sec. V B.

Future research will explore these issues through the analysis of a larger data set in an attempt to estimate the spatial density of mud volcanoes through long-range reverberation measurements and refine the statistical modeling presented in this section.
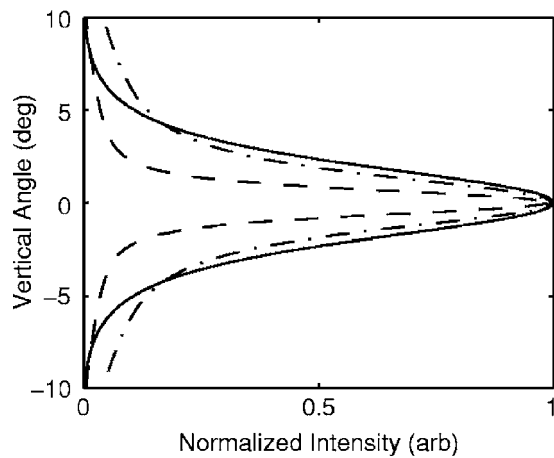
FIG. 13. The distribution of intensity in the vertical plane at 21 km 140 m water depth at 350 (-··-), 800 (---), and 1400 Hz (solid line) assuming isovelocity and range-independent bathymetry. The geoacoustic properties for these predictions were derived from seabed reflection and transmission loss data (see Sec. V B).

FIG. 14. The bathymetry, sound speed profile (gray line), source (∗) and VLA receiver (●) position for the transmission loss measurement in Boundary2002. See Fig. 1 for geographic location.

V. TARGET STRENGTH ANALYSIS

Target strength is a useful quantity that provides a simple measure of the scattered field associated with a feature, in this case a mud volcano. In particular, the target strength is required for the statistical analysis (Sec. IV) and is also of interest to compare with the target strength estimates from direct-path measurements[7] in order to draw conclusions about the similarities or differences between the scattering mechanism between the two time periods. The frequency dependence of the target strength can provide clues about the scattering mechanism, which provides the foundation for development of feature scattering models.

A. Method

The simple sonar equation for target strength (TS) is

$$TS = RL - SL + 2TL, \tag{4}$$

where RL is received level, SL is source level, and TL is the transmission loss. Thus, an estimate of the target strength can be obtained by measuring RL and TL. Here, instead of measuring TL from the source-receiver to the location of the scattering feature (the MV), as required by Eq. (4), we employ nearby TL and seabed reflection measurements to obtain an estimate of the sediment geoacoustic properties and then use those properties to predict TL over the specific source-to-scatterer-to-receiver paths. The source level from the SUS is obtained from empirical relations.[27]

For extended targets, Ref. 28 points out that the sonar equation is only valid under the condition where the scattering cross section is approximately independent of vertical angles $\pm \Delta \theta$ that span the dominant waveguide modes. The vertical angles that are expected at the mud volcanoes as a function of frequency are shown in Fig. 13. This is approximately met for objects of spatial extent L and wavelength λ when $2\Delta\theta < \lambda/(2L)$ or when the object is a sphere "or certain other rounded objects." While the first condition is not met at frequencies above ~500 Hz, the "certain rounded objects condition" is met inasmuch as direct path measure-

ments (800–3600 Hz) indicate that the scattering is nearly independent of vertical angle from near horizontal to 45° grazing (see Ref. 7). Thus, the sonar equation approach should be valid over the entire frequency range of interest here.
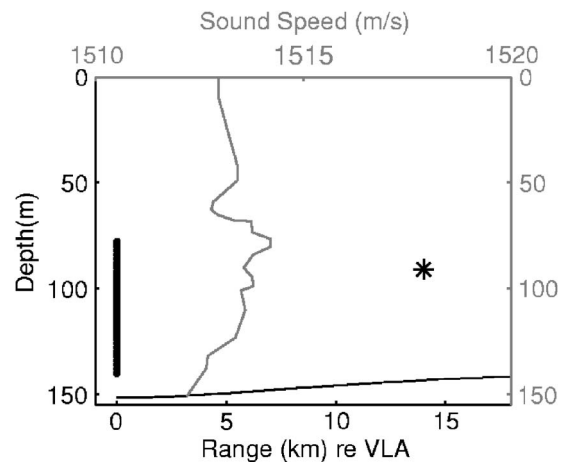
One of the issues in target strength modeling is the proper definition of the incident field on the scatterer. If the scatterer were flush with the seabed, the conventional approach is to assume that the field incident on the scatterer is one-half the full field at the seabed. If the scatterer is in the water column the full field is used. In this case, given our understanding from the direct path measurements,[7] the scattering mechanism appears to be the carbonate chimneys on/near the top of the mud volcanoes (i.e., in the water column). Thus we take the full field. If one-half the full field was correct we would need to add $10 \log_{10}(2^2) = 6$ dB to our target strength estimates. Another issue related to the incident field is the selection of what height above the seabed to take it: toward the base, the middle, or the top of the mud volcano. The correct approach depends again on where the scattering is coming from—the evidence from the direct path results indicate that the scattering is not coming from near the base, the evidence rather points to the carbonate chimneys along the top of the MVs. Thus we use the computed field (TL) from the top of the MV, i.e., 5 m above the seabed.

B. Geoacoustic model development

In order to obtain estimates of the seabed properties we use a combination of transmission loss and seabed reflection data.

Transmission loss data were collected in close proximity to the reverberation measurements (see Fig. 1 for location) during the Boundary2002 Experiment.[11] A series of Mk82 SUS sources were employed at 91 m depth and the receiver was a moored vertical array of 64 elements spanning the lower portion of the water column (see Fig. 14). The pressure time series at 90 m depth and 17.8 km is shown in Fig. 15; total time spread is about 500 ms. Note the three distinct arrivals separated by about 40 ms, which are associated with the bubble pulse. The time series data were processed in 1/3

Holland *et al.*: Long-range scattering from mud volcano cluster
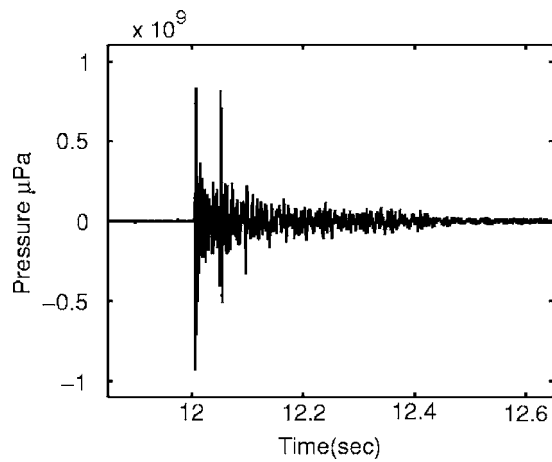
FIG. 15. Received pressure time series at 17.8 km, 90 m depth. Note the arrivals at about 40 ms spacing which are associated with the SUS bubble pulse. That is, for a 0.82 kg source at 91 m, the bubble pulse is predicted to be 42 ms.

TABLE I. Geoacoustic model for area between Sites C1, C5 and the mud volcano cluster.

Thickness (m)	Sound speed (m/s)	Attenuation (dB/m/kHz)	Density (g/cm^3)
1.74	1488	0.05	1.51
2.94	1551	0.1	1.79
1.68	1624	0.1	2
	1563	0.1	1.9

octave frequency bands using source levels from empirical relations[27] to obtain transmission loss (TL), Fig. 16. The field is nearly independent of depth as expected from a nearly isovelocity sound speed profile. Though there is small slope along the track (about 0.03°), the propagation model (PE) shows that this range dependence has a negligible impact on the TL.

Broadband seabed reflection data were collected at the northern end of the propagation track, using a seismic Uniboom source and a single hydrophone receiver (see Ref. 29 for more details on the data collection and processing details). These data were analyzed in the time and frequency domains to obtain a geoacoustic model (see Table I) using a Bayesian approach.[30] The Bayesian approach yielded values for density, sound speed, and layer thickness with modest uncertainties, however, the uncertainties associated with the attenuation were quite large (see Ref. 30). The layer attenuations (in Table I) used in this study were obtained by adjusting the values until a reasonable fit to the TL was obtained. There was no attempt to determine a depth-dependent attenuation profile, other than to allow the first layer, a silty-clay fabric that typically has relatively low attenuation, to have a value distinct from the other layers. The predicted TL using this geoacoustic model compares favorably with the measured TL data in Fig. 16.

C. Mud-volcano target strength results

Target strengths are estimated using Eq. (4) for five shots that occurred within 3 h and ~2 km of each other from Site C1. The received level is obtained by picking off the peak value of the reverberation associated with the time/bearing corresponding to the largest MV in the cluster. Generally, the highest amplitude associated with the reverbera-
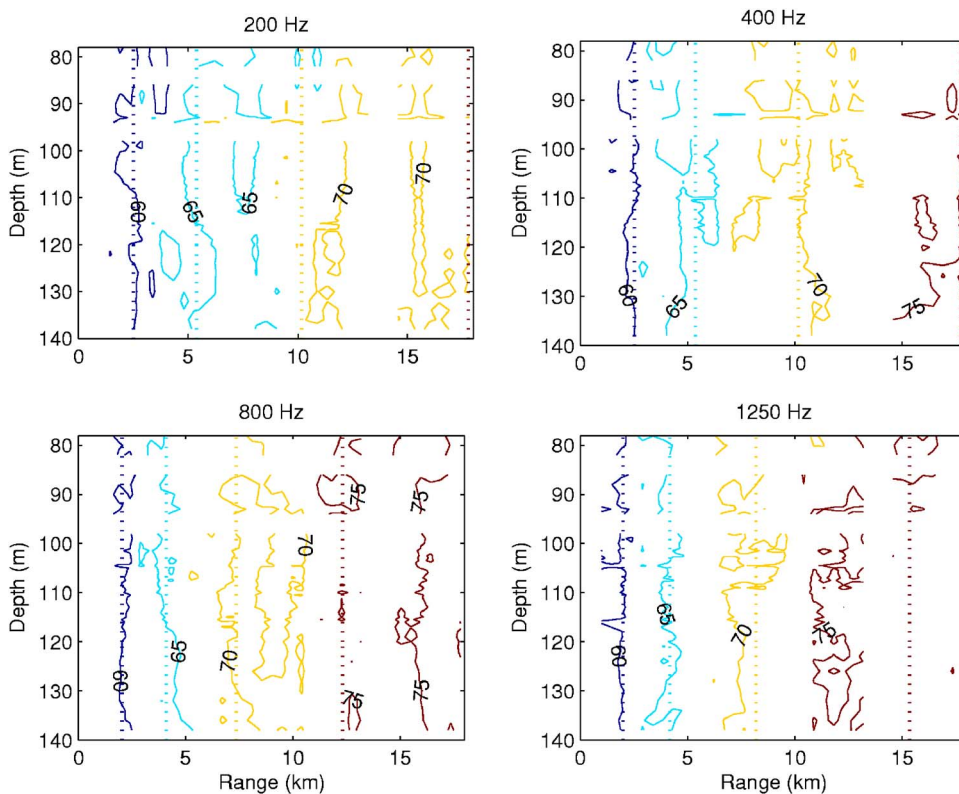
FIG. 16. (Color online) Measured (solid lines) and modeled (dotted) transmission loss (see Fig. 1 for location).
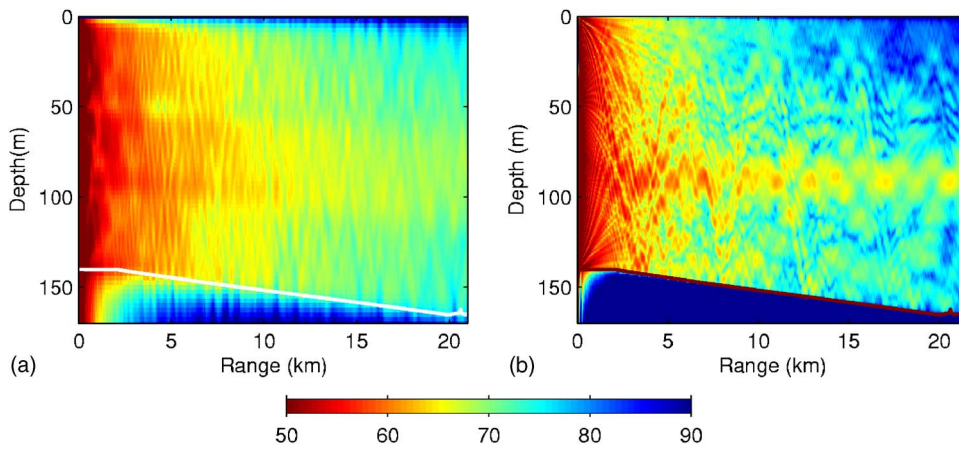
FIG. 17. (Color online) Predicted transmission loss in decibels from Site C1 to MV at (a) 160 Hz and (b) 1400 Hz predicted by PE (RAM) in a 100 Hz band used to estimate target strength. The MV cluster of interest is at about 21 km. The sound speed profiles are shown in Fig. 3(a).

tion from the entire MV field appears to be associated with the largest MV [see Fig. 2(a) at 36.5716N 14.4383E].

The source level from the SUS is obtained from empirical relations,[27] corrected to power spectral density (τ =1/12 s). The two-way TL is estimated using the geoacoustic model in Table I in the PE model RAM including water column attenuation as a simple correction, $e^{-2\beta r}$, where β is the seawater attenuation. The model predictions were averaged incoherently across a 100 Hz band. The bandwidth was selected to be consistent with the statistical analysis (Sec. IV) and is a compromise between narrow bandwidths which result in large fluctuations in the TL, and large bandwidths which reduce the ability to probe the frequency dependence of the target strength. The predicted transmission loss for the Site C1 experiment at 160 and 1400 Hz is shown in Fig. 17. The TL for the target strength is taken at 5 m above the seabed at the range/time associated with the distance to the largest MV for each ping, averaged over the radial dimension ($c\tau/2$) of the scattering area.

The mean target strengths measured from the reverberation are shown in Fig. 18 (open circle); the error bars show the standard deviations for the five shots (for the lower two

frequencies there were only four shots). Also shown are measurements of short-range direct path mean target strength (closed diamond) of the largest MV at low grazing angles 3°–5° (see Ref. 7 for discussion of these data). A few aspects of the two data sets should be noted. First, the inherent variability of the TS is generally much greater for the long-range data than for the direct path data as expected (though admittedly the sample size for the long-range data, four to five shots, is quite small; the short-range data are averaged over 15 pings). In addition, we expect that the inherent accuracy is greater for the direct path data, since no knowledge about the sediment geoacoustic properties is required, and the relative distance/time scales associated with the water column variability are much smaller. Finally, note that the cross-range dimension of the insonified patch for the long-range data is 475 m at 1400 Hz, and is inversely proportional to frequency so that at 160 Hz, the cross-range dimension is 4 km. For the direct path TS data, the source and receiver were omnidirectional in the horizontal plane.

A salient point in this comparison (Fig. 18) is that the target strengths from the two measurement techniques are remarkably similar. We reason that this agreement between the two measurements conducted at two different times indicates that the scattering mechanisms between the two time periods are also similar. Since there is strong evidence that the scattering mechanism in the short-range data was due to carbonate chimneys,[7] it follows that the most likely scattering mechanism for the long-range data is due to the same mechanism. Other possible explanations are briefly explored. It should be noted that the azimuths of the short and long-range measurements differed somewhat, respectively ~35° and 75° relative to the axis of the MV. Thus, another possible explanation for the similar TS is that there were different scattering mechanisms during the two time periods, but the resulting TS difference was precisely compensated for by the azimuthal dependence between the two mechanisms. However, given the relatively modest differences in azimuth (40°), and the fact that the carbonate chimneys appear to be circular in the azimuthal plane so that scattering should be independent of azimuth, this explanation seems unlikely. It might also be argued that the scattering mechanisms might differ, but the target strength remains the same. However,

FIG. 18. Target strength measured from the largest MV using long-range reverberation at Site C1 (○) for a 100 Hz bandwidth. Also shown are measured target strengths (200 Hz bandwidth) from direct-path measurements (♦) from the same MV (Ref. 7). Predicted target strength from a sphere of radius 5 m is shown in the gray line. Above 100 Hz, the model prediction is averaged over a 100 Hz band and above 1600 Hz, the model is averaged over a 200 Hz band (commensurate with the direct-path data).

TABLE II. Properties of an idealized carbonate chimney used to model scattering from a viscoelastic sphere.

Compressional speed (m/s)	Shear speed (m/s)	Compressional attenuation (dB/m/kHz)	Shear attenuation (dB/m/kHz)	Density (g/cm^3)
2200	1000	0.1	20	2.2

invoking Occam's razor, the simplest explanation of the observed target strength consistency is the consistency in the underlying scattering mechanism.

The frequency dependence of the target strength can provide clues as to a suitable model for the scattering. Note that over the decade of frequency, the scattering is roughly independent of frequency. This characteristic suggests using the model of sphere, which gives a scattering strength independent of frequency, for frequencies above $ka \gg 1$, where a is the sphere radius, or characteristic length of the scattering object. Taking the characteristic length as the radius of one of the larger carbonate chimneys, $a = 5$ m, and treating the sphere as viscoelastic, the predicted target strength[31] is in reasonable agreement with the observations (see gray line Fig. 18).

The current measurements extend the prior observations (reported in Ref. 7) down by several octaves in frequency. Since there is evidence of multiple carbonate chimneys on some of the MVs, it is expected that the total target strength is generally a sum of several carbonate chimneys, the number depending on the size of the insonified area. Assuming a relatively small number of carbonate chimneys insonified in a given patch (side-scan data[7] suggest that the total number is small, approximately less than 5 for the data here) $a = 2$ m is a reasonable lower bound to the largest characteristic scale. In other words, if the largest characteristic scale was much smaller than $a = 2$ m, the predicted target strength at 160 Hz would be much lower than observed. The geoacoustic parameters for the proxy sphere model are given in Table II. Though the actual properties of the carbonate chimneys are unknown, it should be noted that the scattering strength is relatively insensitive to the geoacoustic properties (see the discussion in Ref. 7).

VI. DISCUSSION AND CONCLUSIONS

Analysis of reverberation measurements in a littoral region in the Straits of Sicily show strong returns associated with mud volcanoes. This fact leads to both a challenge and an opportunity. The challenge is that the presence of mud volcanoes in littoral regions likely leads to performance degradation for active sonars, due to increased probability of false alarm. The statistical analysis confirms that the reverberation associated with the mud volcanoes is strongly non-Rayleigh.

On the other hand, the global climate change community has raised the hypothesis that shallow-water mud volcanoes may be a significant and, as yet, missing contributor to the overall atmospheric methane budget.[32] Measurements near the mud volcanoes show substantially elevated levels of methane.[7] Since the scale of shallow water mud volcanoes is

fairly small, it means that traditional geophysical survey methods for finding them will be costly. The opportunity for the global climate community is that using an active sonar approach will increase survey coverage rates by two to three orders of magnitude.

Gas ebullition from mud volcanoes is inherently episodic and the time scales of ebullition are expected to vary widely, depending on the local geologic and oceanographic characteristics. It is believed that the mud volcanoes studied here are "active," the evidence for this being that there are no bio-constructions[7] on the edifice. However, during visual and sonar observations in 2004, there was no evidence of bubbles from the mud volcanoes (though there were high levels of dissolved methane seeping from the seabed). In addition, the direct path measurements showed that scattering from the water column above the mud volcanoes was negligible.[7] The similarity of the targets strengths from the 2000 data thus suggest that the scattering was caused by the same mechanism (i.e., not bubbles). Given so few observations, it is not possible to infer anything about the time scales of ebullition for these particular mud volcanoes. Knowledge of the time scales would be quite important, however, inasmuch as the reverberation statistics are expected to be very different depending on whether or not gas ebullition (in the form of emitted bubbles) is present. The leading hypothesis for the scattering mechanism in the 100–5000 Hz band is the carbonate chimneys (of order 10 m in diameter and several meters in height) on the mud volcanoes. Thus, the results presented here represent our understanding of acoustic scattering from mud volcanoes in a quiescent state.

The results here are also valuable for developing clutter models. A statistical model comprising both K- and Rayleigh-distributed components may be used to examine sonar performance or simulate data from mud volcanoes where the mud-volcano shape parameter is taken as constant and quite small, roughly 0.1 to 0.2. The scattering amplitude of the mud volcanoes can be modeled by a sphere of comparable dimension to the carbonate chimneys.

ACKNOWLEDGMENTS

[1] M. K. Prior, "A scatterer map for the Malta Plateau," IEEE J. Ocean. Eng. **30**, 676–690 (2005).
[2] C. W. Holland, "Mapping seabed variability: Rapid surveying of coastal regions," J. Acoust. Soc. Am. **119**, 1373–1387 (2006).
[3] P. Ratilal et al., "Long-range acoustic imaging of the continental shelf environment: The acoustic clutter reconnaisance experiment 2001," J.

Acoust. Soc. Am. **117**, 1977–1998 (2005).

[4]N. C. Makris, C. S. Chia, and L. T. Fialkowski, "The biazimuthal scattering distribution of an abyssal hill," J. Acoust. Soc. Am. **106**, 2491–2512 (1999).

[5]A. V. Milkov, "Worldwide distribution of submarine mud volcanoes and associated gas hydrates," Mar. Geol. **167**, 29–42 (2000).

[6]M. B. Cita, J. M. Woodside, M. K. Ivanovov, R. B. Kidd, A. F. Limonov, and Scientific Staff of Cruise TTR3-Leg 2, "Fluid venting from a mud volcano on the Mediterranean Ridge," Terra Nova **7**, 453–458 (1995).

[7]C. W. Holland, T. Weber, and G. Etiope, "Acoustic scattering from mud volcanoes and carbonate mounds," J. Acoust. Soc. Am. **120**, 3553–3565 (2006).

[8]In the short-range measurement, the vessel drifts with the vertical array deployed a few tens of meters above the bottom, the direction of the drift being governed by wind and currents.

[9]M. Grasso and C. D. Reuther, "The western margin of the Hyblean Plateau: A neotectonic transform system on the SE Sicilian foreland," Annales Tectonicae **2**, 107–120 (1988).

[10]C. W. Holland, G. Etiope, A. V. Milkov, E. Michelozzi, and P. Favali, "Mud volcanoes discovered offshore Sicily," Mar. Geol. **199**, 1–6 (2003).

[11]C. W. Holland *et al.*, "Boundary characterization experiment series overview," IEEE J. Ocean. Eng. **30**, 784–806 (2005).

[12]J. Preston, D. Ellis, and R. Gauss, "Geoacoustic parameter extraction using reverberation data from the 2000 Boundary Characterization experiment on the Malta Plateau," IEEE J. Ocean. Eng. **30**, 709–732 (2005).

[13]The four shots were selected based on high SNR, which means geometries in which the beams were not too close to endfire and beams that were not contaminated by ship noise.

[14]D. A. Abraham and P. K. Willett, "Active sonar detection in shallow water using the Page test," IEEE J. Ocean. Eng. **27**, 35–46 (2002).

[15]Y. Y. Dorfman and I. Dyer, "Monostatic and bistatic reverberation statistics west of the mid-Atlantic ridge," J. Acoust. Soc. Am. **106**, 1755–1764 (1999).

[16]W. D. Hibler III, W. F. Weeks, and S. J. Mock, "Statistical aspects of sea-ice ridge distributions," J. Geophys. Res. **77**, 5954–5970 (1972).

[17]P. Wadhams, "Sea-ice topography of the Arctic ocean in the region 70°W to 25°E," Proc. R. Soc. London, Ser. A **302**, 45–85 (1981).

[18]K. D. LePage, "Statistics of broad-band bottom reverberation predictions in shallow-water waveguides," IEEE J. Ocean. Eng. **29**, 330–346 (2004).

[19]K. D. LePage, "Modeling higher moments of reverberation in shallow water waveguides," in Proceedings of the Eighth European Conference on Underwater Acoustics, edited by S. M. Jesus and O. C. Rodriguez, Carvoeiro, Portugal, June 2006, pp. 473–478.

[20]E. Jakeman and R. J. A. Tough, "Non-Gaussian models for the statistics of scattered waves," Adv. Phys. **37**, 471–529 (1988).

[21]D. A. Abraham and A. P. Lyons, "Novel physical interpretations of K-distributed reverberation," IEEE J. Ocean. Eng. **27**, 800–813 (2002).

[22]M. Fisz, *Probability Theory and Mathematical Statistics*, 3rd ed. (Krieger, Malabar, FL, 1963).

[23]D. A. Abraham and M. K. Prior, "Estimating independent-scatterer density from active sonar reverberation," in Proceedings of Eighth European Conference on Underwater Acoustics, Carvoeiro, Portugal, June 2006, pp. 181–186.

[24]D. A. Abraham and A. P. Lyons, "Bootstrapped K-distribution parameter estimation," in Proceedings of IEEE/MTS Oceans 2006 Boston Conference, Boston, MA, 2006, pp. 1–6.

[25]J. R. Preston and D. A. Abraham, "Non-Rayleigh reverberation characteristics near 400 Hz observed on the New Jersey Shelf," IEEE J. Ocean. Eng. **29**, 215–235 (2004).

[26]D. A. Abraham, A. P. Lyons, and K. M. Becker, "The effect of multipath on reverberation envelope statistics," Proceedings of the Seventh European Conference on Underwater Acoustics, Delft, The Netherlands, 2004, pp. 295–300.

[27]N. R. Chapman, "Source levels of shallow explosive charges," J. Acoust. Soc. Am. **84**, 697–702 (1988).

[28]P. Ratilal, Y. Lai, and N. Makris, "Validity of the sonar equation and Babinet's principle for object scattering in a shallow water waveguide," J. Acoust. Soc. Am. **112**, 1797–1816 (2002).

[29]C. W. Holland, "Seabed reflection measurement uncertainty," J. Acoust. Soc. Am. **114**, 1861–1873 (2003).

[30]J. Dettmer, S. E. Dosso, and C. W. Holland, "Bayesian inversion of broadband reflection data for seabed properties of multi-layered systems," Proceedings of the Eighth European Conference on Underwater Acoustics, edited by S. M. Jesus and O. C. Rodriguez, Carvoeiro, Portugal, June 2006, pp. 473–478.

[31]J. F. Faran, "Sound scattering by solid cylinders and spheres," J. Acoust. Soc. Am. **23**, 405–418 (1951).

[32]G. Etiope and A. V. Milkov, "A new estimate of global methane flux from onshore and shallow submarine mud volcanoes to the atmosphere," Environ. Geol. **46**, 997–1002 (2004).

A statistical geoacoustic inversion scheme based on a modified radial basis functions neural network

George Tzagkarakis[a)]
*Department of Computer Science, University of Crete and Institute of Computer Science – FO.R.T.H.
FORTH-ICS, P.O. Box 1385, 711 10 Heraklion, Crete, Greece*

Michael I. Taroudakis[b)]
*Department of Mathematics, University of Crete and Institute of Applied and Computational Mathematics –
FO.R.T.H. FORTH-IACM, P.O. Box 1385, 711 10 Heraklion, Crete, Greece*

Panagiotis Tsakalides[c)]
*Department of Computer Science, University of Crete and Institute of Computer Science – FO.R.T.H.
FORTH-ICS, P.O. Box 1385, 711 10 Heraklion, Crete, Greece*

This paper addresses the task of recovering the geoacoustic parameters of a shallow-water environment using measurements of the acoustic field due to a known source and a neural network based inversion process. First, a novel efficient "observable" of the acoustic signal is proposed, which represents the signal in accordance with the recoverable parameters. Motivated by recent studies in non-Gaussian statistical theory, the observable is defined as a set of estimated model parameters of the alpha-stable distributions, which fit the marginal statistics of the wavelet subband coefficients, obtained after the transformation of the original signal via a one-dimensional wavelet decomposition. Following the modeling process to extract the observables as features, a radial basis functions neural network is employed to approximate the vector function that takes as input the observables and gives as output the corresponding set of environmental parameters. The performance of the proposed approach in recovering the sound speed and density in the substrate of a typical shallow-water environment is evaluated using a database of synthetic acoustic signals, generated by means of a normal-mode acoustic propagation algorithm.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2772232]

I. INTRODUCTION

Inverse problems in underwater acoustics are associated with measurements of the acoustic field performed in the frequency or in the time domain. In this framework, a set **d** of observables is defined, which forms the input parameters of the inverse problem. The observables are related to the recoverable environmental parameters **m** through a linear or nonlinear vector equation of the form $\mathcal{T}(\mathbf{d},\mathbf{m})=0$.

The inversion procedure is based on the properties of the relationship between **m** and **d**, and it is considered to be more efficient if even small variations of the environmental parameters are associated with observables which can be clearly discriminated via the mapping $\mathcal{T}(\cdot,\cdot)$. Since the performance of a specific inversion procedure is directly related to the selected observables, defining observables which are easily identifiable and as sensitive as possible to changes of the environmental parameters, constitutes an important task.

Determining the sea-bed parameters from acoustic measurements obtained in the water column is among the most interesting inverse problems in underwater acoustics.[1–4] It should be noted that most of the inversion procedures and the associated observable identification (feature extraction) are based on deterministic approaches. In recent work,[5,6] we proposed a novel observable for acoustic signals, based on a *symmetric alpha-stable* (*SαS*) statistical modeling of the coefficients obtained after a transformation of the original signal using the one-dimensional (1D) dyadic wavelet transform (DWT). Then, a classification scheme was designed by combining the extracted features, that is, the estimated *SαS* parameters at each wavelet subband, with a closed-form expression of the Kullback-Leibler divergence (KLD) between *SαS* distributions. First results based on synthetic data showed that the proposed scheme provided a very accurate classification of a recorded acoustic signal in the true unknown environment, specified by several sets of parameters (e.g., sound speed profiles in the water and/or bottom domains, layer thicknesses and densities, source location, etc.).

Based on the above, in this paper we treat the inverse problem as a function approximation problem. In particular, we consider a nonlinear mapping $\mathcal{T}$ with arguments the estimated *SαS* parameters and output the set of the corresponding environmental parameters. Our goal is to find an accurate approximation of $\mathcal{T}$. An efficient approximation of such a mapping between an acoustic field and its corresponding geoacoustic parameters is achieved using neural network-based approaches. In previous studies,[7,8] the inversion process was carried out by employing multilayer feed-forward

---
a)Electronic mail: gtzag@ics.forth.gr
b)Electronic mail: taroud@iacm.forth.gr
c)Electronic mail: tsakalid@ics.forth.gr

neural networks. In the present work, a neural network-based approach is employed using radial basis functions (RBF), since it is well known that RBF neural networks provide a suitable approximation to nonlinear functions in an efficient way. In particular, we propose a modified version of the standard RBF neural network, by replacing the Euclidean distance function, which measures the similarity between the input vectors and the centers of the hidden neurons, with the KLD between $S\alpha S$ distributions.

The paper is organized as follows: In Sec. II, the statistical modeling via $S\alpha S$ distributions is described, and the procedure for the design of the modified RBF network based on the KLD between $S\alpha S$ distributions is analyzed in detail. In Sec. III, the proposed RBF neural network is applied to a database of simulated acoustic signals generated in a shallow water environment to evaluate the inversion performance. Finally, we conclude in Sec. IV giving some future research directions.

II. THEORETICAL BACKGROUND

In the following, we introduce the main mathematical background of the proposed inversion scheme.

A. Statistical modeling via $S\alpha S$ distributions

The proposed geoacoustic inversion process is performed in two steps: First, a feature extraction procedure is applied, which represents the information content of a given acoustic signal with an appropriate set of features and second, the extracted features are used to build a RBF neural network.

During the feature extraction step, the acoustic signal is decomposed into several levels through a multiresolution analysis employing the 1-D DWT. This transform works as follows: Starting from the given signal $s(t)$, two sets of coefficients are computed at the first level of decomposition, (i) approximation coefficients A1 and (ii) detail coefficients D1. These vectors are obtained by convolving $s(t)$ with a low-pass filter for approximation and with a high-pass filter for detail, followed by dyadic decimation. At the second level of decomposition, the vector A1 of the approximation coefficients is decomposed in two sets of coefficients using the same approach replacing $s(t)$ by A1 and producing A2 and D2. This procedure continues in the same way, namely at the kth level of decomposition we filter the vector of the approximation coefficients computed at the $(k-1)$th level. Thus, when $s(t)$ is decomposed in L levels, we obtain a set of L detail subbands (containing the high-frequency content) and one approximation subband (containing the low-frequency content).

For short-time, pulse shallow-water acoustic signals, the 1-D DWT seems to be a powerful modeling tool, providing a natural arrangement of the wavelet coefficients into multiple levels representing the frequency content of the signal in consecutive bands.[9] Besides, it has been pointed out that the wavelet transforms of signals which present such a transient behavior tend to be sparse, resulting in a large number of coefficients with small magnitude and a small number of large magnitude coefficients.[10] This property gives rise to peaky and heavy-tailed *non-Gaussian* marginal distributions of the wavelet subband coefficients.[10]

Following the acoustic signal decomposition, an accurate fitting of the tails of the marginal distribution of the wavelet coefficients at each subband is achieved by modeling them as $S\alpha S$ random variables. A $S\alpha S$ distribution is best defined by its characteristic function:[11]

$$\phi(t) = \exp(i\delta t - \gamma^\alpha |t|^\alpha), \tag{1}$$

where α is the *characteristic exponent*, taking values $0 < \alpha \leq 2$, $\delta(-\infty < \delta < \infty)$ is the *location parameter*, and $\gamma(\gamma > 0)$ is the *dispersion* of the distribution. The characteristic exponent is a shape parameter, which controls the "thickness" of the tails of the density function. The smaller the α, the heavier the tails of the $S\alpha S$ density function. The dispersion parameter determines the spread of the distribution around its location parameter, similar to the variance of the Gaussian.

In our previous work,[5] the content of an underwater signal was accurately represented with a set of features, which is much smaller in size than the signal itself or any other representation in the frequency or wavelet domain. This set contains the maximum likelihood (ML) estimated parameters (α, γ) of the $S\alpha S$ distribution at each wavelet subband.

Thus, for a given acoustic signal S, decomposed in L levels, its feature vector $\mathbf{d}$ is given by the following set of $L+1$ pairs:

$$S \mapsto \mathbf{d} = \{(\alpha_1, \gamma_1), (\alpha_2, \gamma_2), \ldots, (\alpha_{L+1}, \gamma_{L+1})\}, \tag{2}$$

where (α_i, γ_i) are the estimated model parameters of the ith subband, using the consistent ML method described by Nolan,[12] which gives reliable estimates and provides the tightest confidence intervals. Note also that we follow the convention that $i=1$ corresponds to the detail subband at the first decomposition level, while $i=L+1$ corresponds to the approximation subband at the Lth level.

B. Inversion using a RBF neural network

As we demonstrated in a recent work,[5] the KLD[13] is capable of distinguishing between two distinct acoustic signals, since the KLD between two signals generated in similar shallow-water environments is almost zero, whereas it increases when the signals are obtained from different environments. Thus, the function $\mathcal{T}$, which maps the estimated $S\alpha S$ parameters $\mathbf{d}$ to the corresponding environmental parameters $\mathbf{m}$, is a well-defined nonlinear vector function. We also assume that $\mathcal{T}$ defines a one-to-one correspondence between $\mathbf{d}$ and $\mathbf{m}$.

This observation yields that there is a nonlinear vector function $\mathcal{T}: \mathcal{A} \mapsto \mathbb{R}^n$, where $\mathcal{A} \subseteq \mathbb{R}^{2(L+1)}$ contains the estimated parameters $\{(\alpha_1, \gamma_1), \ldots, (\alpha_{L+1}, \gamma_{L+1})\}$ of each signal and n is the number of the environmental parameters we are interested in. In this paper, the shallow-water environment is modeled as a two-layered medium, with the first layer representing the water column and the second semi-infinite layer representing the substrate. Then, we focus on the recovery of the sound speed in the substrate, c_{sb}, and the substrate density ρ_{sb}. That is, the function $\mathcal{T}$ maps the $S\alpha S$ parameters in

Tzagkarakis *et al.*: Statistical geoacoustic inversion via neural networks

FIG. 1. The radial basis neuron model.

FIG. 2. The general architecture of a RBF network.
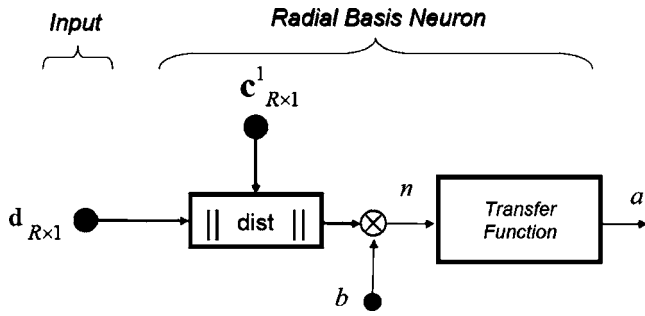
$\mathbb{R}^2$, and particularly in vectors of the form $\mathbf{m}=[c_{\rm sb},\rho_{\rm sb}]$. Note that a similar problem can be defined for each set of recoverable environmental parameters. In our previous work,[5] it was shown that the KLD is capable of performing the identification of the parameters characterizing the sound speed profile in the water column as well.

Working in this framework, we proceed by converting the inverse problem into a function approximation problem. In particular, if we were able to approximate accurately the function $\mathcal{T}$, then we would be able to find a solution to the inverse problem, since the insertion of the estimated $S\alpha S$ parameters of a signal, recorded in an unknown environment, into $\mathcal{T}$, would result in the computation of the vector $\mathbf{m}$ $=[c_{\rm sb},\rho_{\rm sb}]$ with elements the substrate parameters of the unknown environment.

An efficient process to solve the above-mentioned function approximation problem is achieved by using a RBF neural network,[14] as it is well known that such a network is suitable for the approximation of a nonlinear function. A RBF network consists of two layers: a hidden radial basis layer of N_1 neurons, and an output linear layer of N_2 neurons. In the test case described in Sec. II C, N_1 can be at most equal to the number of training samples (M) and $N_2=2$, since we are interested in recovering the two environmental parameters $[c_{\rm sb},\rho_{\rm sb}]$.

Figure 1 shows the model of a single radial basis neuron in the hidden layer. In particular, the net input, n, to the transfer function is the vector distance between its center $\mathbf{c}$ and the input vector $\mathbf{d}$, multiplied by the bias b. Thus, the output of the radial basis neuron, a, is equal to the value of the selected transfer function evaluated at $b\cdot\|\mathbf{d}-\mathbf{c}\|$.

Figure 2 shows the general architecture of a RBF network. Each radial basis neuron of the hidden layer is denoted by using its corresponding transfer function ϕ_i, $i=1,\ldots,N_1$, while each neuron of the output linear layer is denoted by L_j, $j=1,\ldots,N_2$. Following the notation of Fig. 1, the output of a single hidden radial basis neuron is given by $a_i=\phi_i(b_i\cdot\|\mathbf{d}-\mathbf{c}_i\|)$, where b_i and $\mathbf{c}_i$ are the bias and the center of the ith radial basis neuron, respectively. Then, the output of a neuron in the linear layer is given by $\mathbf{m}_j=\Sigma_{i=1}^{N_1}w_{i,j}a_i+b_j$. In a common RBF network, the $\|$dist$\|$ box in Fig. 1 measures the Euclidean distance between the input vector $\mathbf{d}$ and the center of the ith neuron, $\mathbf{c}_j$. The transfer function for a radial basis neuron is $\phi(n)=e^{-n^2}$, which has a maximum of 1 when its input is 0. As the distance between $\mathbf{c}_i$ and $\mathbf{d}$ decreases, the output increases. Thus a radial basis neuron acts as a detector
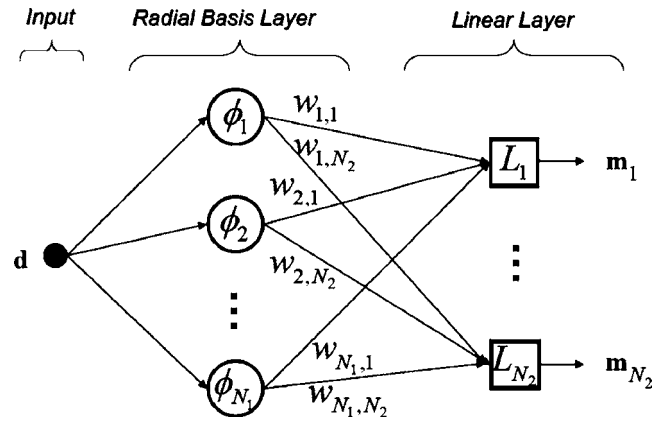
which produces 1 whenever the input $\mathbf{d}$ is identical to its center $\mathbf{c}_i$. Besides, the bias b_i allows the sensitivity of the ith radial basis neuron to be adjusted. For example, if a neuron i had a bias of 0.1 its transfer function $\phi_i(n)$ would output 0.5 for any input vector $\mathbf{d}$ at vector distance of 8.326 ($0.8326/b_i$) from its center $\mathbf{c}_i$.

Thus, the jth component, $\mathbf{m}_j$, of the vector function $\mathcal{T}$ is approximated as a linear combination of the set of radial basis functions:

$$\widetilde{\mathcal{T}}_j(\mathbf{d}) \equiv \widetilde{\mathbf{m}}_j = \sum_{i=1}^{N_1} w_{i,j}\phi_i(\mathbf{d}) + b_j \overset{b_j=w_{0,j}}{=} \sum_{i=0}^{N_1} w_{i,j}\phi_i(\mathbf{d}), \quad (3)$$

where $w_{i,j}$ is equal to the weight of the edge connecting the ith radial basis neuron with the jth output of the network (see Fig. 2). In Eq. (3), the auxiliary radial basis function $\phi_0(\mathbf{d})$ is the constant function $\phi_0(\mathbf{d})=1$. The most common form of basis function used is the Gaussian

$$\phi_i(\mathbf{d}) \equiv \phi_i(b_i\cdot\|\mathbf{d}-\mathbf{c}_i\|) = \exp\left(-\frac{\|\mathbf{d}-\mathbf{c}_i\|^2}{2\sigma_i^2}\right),$$

where $\mathbf{c}_i$ and σ_i are the mean and standard deviation of the basis function, respectively. This basis function is of the form e^{-n^2} with $n=b_i\cdot\|\mathbf{d}-\mathbf{c}_i\|$, where $b_i=1/\sqrt{2}\sigma_i$. A hidden neuron is more sensitive to input vectors near its center. This sensitivity may be tuned by adjusting the widths (spread parameters) σ_i. For a given input vector, typically only a few hidden units will have significant activations. Besides, the spread parameters should be chosen large enough so that neurons respond strongly to overlapping regions of the input space, but they should not be too large so that each neuron would effectively respond in the same large area of the input space.

In the following, we study an inversion scheme based on two different kinds of RBF networks. The first one employs the standard RBF network architecture described so far, that is, the $\|$dist$\|$ in Fig. 1 is the common Euclidean distance and the transfer functions ϕ_i are Gaussians. The second novel scheme is based on a modification of the standard RBF architecture. In particular, we are interested in exploiting the results of our previous work,[5] where it was illustrated that two distinct acoustic signals represented by their corresponding feature vectors can be discriminated very accurately by

employing a version of the KLD between two $S\alpha S$ distributions. Motivated by these results, we modify the standard RBF network by replacing the similarity measure between the input layer and the hidden radial basis layer, namely, by replacing the Euclidean distance with the KLD between $S\alpha S$ distributions.

C. Modified RBF network

In the modified RBF network the similarity measurement between an input vector $\mathbf{d}$ and the ith radial basis neuron's center $\mathbf{c}_i$ is carried out by employing the KLD between two $S\alpha S$ distributions. In our previous study,[5] we showed that a chain rule can be applied in order to combine the KLDs from the multiple wavelet subbands. In particular, the overall distance between two acoustic signals S_1 and S_2, which are represented by the feature vectors $\mathbf{d}_1$ and $\mathbf{d}_2$ given by Eq. (2), respectively, has the following expression:

$$D(S_1,S_2) \equiv D(\mathbf{d}_1,\mathbf{d}_2) = \sum_{k=1}^{L+1} D\ (\hat{q}_{S_1,k}\|\hat{q}_{S_2,k}), \quad (4)$$

where $D(\hat{q}_{S_1,k}\|\hat{q}_{S_2,k})$ is the KLD between the kth wavelet subbands of the two signals, which is evaluated using the estimated $S\alpha S$ parameters $(\alpha_{1,k},\gamma_{1,k})$ and $(\alpha_{2,k},\gamma_{2,k})$, respectively. This marginal KLD is given by

$$D\ (\hat{q}_{S_1,k}\|\hat{q}_{S_2,k}) = \ln\left(\frac{l_{2,k}}{l_{1,k}}\right) - \frac{1}{\alpha_{1,k}}$$
$$+ \left(\frac{\gamma_{2,k}}{\gamma_{1,k}}\right)^{\alpha_{2,k}} \cdot \frac{\Gamma\left(\frac{\alpha_{2,k}+1}{\alpha_{1,k}}\right)}{\Gamma\left(\frac{1}{\alpha_{1,k}}\right)}, \quad (5)$$

with $l_{i,k}$, $i=1,2$, being a normalizing factor, which is equal to

$$l_{i,k} = \frac{2\Gamma\left(\frac{1}{\alpha_{i,k}}\right)}{\alpha_{i,k}\gamma_{i,k}}, \quad (6)$$

where $\Gamma(\cdot)$ is the gamma function.

We modify the standard radial basis neuron model, shown in Fig. 1, by replacing the Euclidean distance (computed by the $\|\text{dist}\|$ box) with the overall KLD between vectors containing estimated $S\alpha S$ parameters (4). Accordingly, the modified output of a single hidden radial basis neuron is given by

$$a_i^{\text{mod}} = \phi_i(b_i \cdot D(\mathbf{d},\mathbf{c}_i)), \quad (7)$$

where $D(\mathbf{d},\mathbf{c}_i)$ is the overall KLD between the input vector $\mathbf{d}$ and the center $\mathbf{c}_i$ of the ith radial basis neuron. The transfer functions $\{\phi_i(\cdot)\}_{i=1,\dots,N_1}$ have the form,

$$\phi_i(\mathbf{d}) \equiv \phi_i(b_i \cdot D(\mathbf{d},\mathbf{c}_i)) = \exp(-b_i \cdot (D(\mathbf{d},\mathbf{c}_i))^2),$$

where again the bias b_i is inversely proportional to the spread of the radial basis function, that is, $b_i \sim 1/\sigma_i$. Note that the KLD is always non-negative, $D(\mathbf{d}_1,\mathbf{d}_2) \geqslant 0$, with an equality if and only if the corresponding $S\alpha S$ parameter pairs are equal, $(\alpha_{1,k},\gamma_{1,k}) = (\alpha_{2,k},\gamma_{2,k})$, $k=1,\dots,L+1$. Thus, the transfer function of the ith neuron has an output equal to 1 if and

only if its center $\mathbf{c}_i$ is equal to the input vector $\mathbf{d}$.

Notice that the KLD between $S\alpha S$ distributions, given by Eq. (4), is not an actual distance function, since it does not satisfy the property of symmetry. However, we can define a symmetrized version of the KLD as follows:

$$D_{\text{sym}}(\mathbf{d}_1,\mathbf{d}_2) = D(\mathbf{d}_1,\mathbf{d}_2) + D(\mathbf{d}_2,\mathbf{d}_1). \quad (8)$$

It can be easily verified that $D_{\text{sym}}(\cdot,\cdot)$ satisfies all the properties of a distance function. In the subsequent illustrations, we will also test the performance of a RBF network based on the symmetrized KLD, that is, whose radial basis transfer functions have the form

$$\phi_i(\mathbf{d}) = \exp(-b_i \cdot (D_{\text{sym}}(\mathbf{d},\mathbf{c}_i))^2).$$

D. Training process

In our proposed scheme, the (standard or modified) RBF network is trained using two different processes, namely, an exact and a more efficient one (let **P1** and **P2** denote the exact and the efficient process, respectively). The exact training process computes the weights $\{w_{i,j}\}_{i=1,\dots,N_1,j=1,\dots,N_2}$ and the biases $\{b_j\}_{j=1,\dots,N_2}$, such that the produced network achieves zero error on the training vectors. Besides, the exact process creates as many radial basis neurons as there are input vectors, where the ith input vector is used as the center $\mathbf{c}_i$ of the corresponding neuron. The drawback of the exact process is that it produces a large network when many input vectors are needed to properly define a network.

On the other hand, the second and more efficient training process produces the RBF network iteratively, by adding one neuron at a time, starting with a single neuron. At each iteration, the input vector that will result in lowering the network error the most is used to create a radial basis neuron. The error of the new network is checked, and if it is low enough the training process terminates. Otherwise, the next neuron is added. Neurons are added to the network until the sum-squared error (SSE) falls below a specified error threshold or a maximum number of neurons is reached. Given a network with K radial basis neurons and input-output pairs $\{(\mathbf{d}_k,\mathbf{m}_k)\}_{k=1,\dots,K}$, the SSE is given by

$$\text{SSE} = \sum_{k=1}^{K} \|\tilde{\mathcal{T}}(\mathbf{d}_k) - \mathbf{m}_k\|^2, \quad (9)$$

where $\tilde{\mathcal{T}}(\mathbf{d}_k)$ is the approximation of the function $\mathcal{T}$ at the input point $\mathbf{d}_k$, obtained at the output of the RBF network.

In both of the above-presented training processes it is important that the spread parameter of the transfer function of each radial basis neuron be large enough so that the neurons respond to overlapping regions of the input space. This also results in a better generalization for new input vectors occurring between input vectors used in the training process. However, the spread parameter should not be too large that all the neurons respond in essentially the same manner.

For the standard transfer function, $\phi(n)=e^{-n^2}$, the bias, b_i, of the ith radial basis neuron is related to the spread, σ_i, of the corresponding transfer function with the expression: $b_i = \sqrt{-\ln(0.5)}/\sigma_i$. This means that if the neuron's center $\mathbf{c}_i$ is at

TABLE I. The shallow water environment.

Water Depth (H)	200 m
Range (R)	5 km
Central frequency (f_0)	100 Hz
Bandwidth (Δf)	40 Hz
Source/receiver depth	100 m
Sound speed profile in the water:	
$c_w(0)$	1500 m/s
$c_w(\min)$	1490 m/s
$c_w(H)$	1515 m/s
d [depth of min $c_w(z)$]	50 m
Semi-infinite substrate: (varying parameters)	
c_{sb}	[1550, 1650] m/s
ρ_{sb}	[1170, 1240] kg/m^3

FIG. 3. The shallow water sea environment used in the present study.

a distance of σ_i from the input vector $\mathbf{d}$, then the output of the transfer function will be 0.5. In the subsequent illustrations, the influence of the exponent value of the transfer function is also studied. In particular, we design RBF networks whose transfer function has the general form $\phi(n) = e^{-n^r}$, where r is a positive integer, which is also known as a *sharpness* parameter. In this case, the bias is related to the spread parameter via the expression: $b_i = \sqrt[r]{-\ln(0.5)}/\sigma_i$.

A common heuristic for the selection of the spread parameter of a radial basis neuron is the following: Choose a spread constant larger than the distance between adjacent input vectors, so as to get good generalization, but smaller than the distance across the whole input space. Thus, in the standard RBF network it should be ensured that $d_{E,\min} \leq \sigma_i \leq d_{E,\max}$, where $d_{E,\min}$ and $d_{E,\max}$ are the minimum and maximum Euclidean distances between the training input vectors, respectively. Similarly, in the modified RBF network one should ensure that $d_{KLD,\min} \leq \sigma_i \leq d_{KLD,\max}$, where $d_{KLD,\min}$ and $d_{KLD,\max}$ are the minimum and maximum KLDs between the entraining input vectors, respectively, given by Eq. (4) or (8).

III. APPLICATION OF THE INVERSION SCHEME USING SYNTHETIC DATA

In this section, the efficiency of the proposed inversion scheme for shallow-water acoustic transmissions is evaluated using synthetic signals, based on the range independent and axially symmetric environment described in Table I. Figure 3 shows the sea environment of the experimental setup consisting of a shallow-water layer and a semi-infinite bottom (the substrate), which are considered fluid. The sound speed profile may vary with depth in the water layer, while it is constant in the substrate. Here, a linear sound speed profile in the water column is considered. For simplicity, the density of both layers is assumed to be constant.

As an attempt to simulate a geoacoustic inversion experiment, we consider a low-frequency sound source, with central frequency $f_0 = 100$ Hz and bandwidth $\Delta f = 40$ Hz. The source excitation function is modeled as a Gaussian, placed at a known depth of 100 m. A single receiver is placed at a distance of 5 km from the source and at the same depth. The
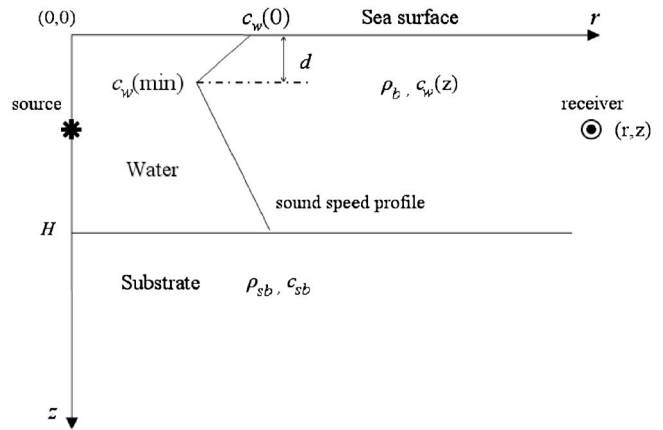
experiment is aiming at the identification of the sea bed, characterized by the sound speed and density of the substrate.

We constructed a database by generating a set of synthetic signals according to the environmental parameters shown in Table I. The sound speed in the substrate varies in the interval [1550, 1650] m/s with a step size equal to 5 m/s and the substrate density varies in the interval [1170, 1240] kg/m^3 with a step size equal to 1 kg/m^3, resulting in a set with a total of 1491 synthetic acoustic signals. The signals are calculated using the Normal-Mode program MODE1 developed at FO.R.T.H-I.A.C.M. Then, the obtained data are provided as input to the inverse discrete Fourier transform to yield the signals in the time domain. Each of the time-domain signals is decomposed by implementing a three-level 1D DWT using the db4 wavelet function and, thus, each signal is represented by a vector $\mathbf{d}$ with $2(L+1) = 8$ elements given by Eq. (2). The experimental results in our recent work[5] showed that the best classification performance is obtained for the db4 wavelet function and using the estimated $S\alpha S$ parameters of the detail subbands only. Thus, in the subsequent illustrations each signal is represented by a vector $\mathbf{d}$ with $2L=6$ parameters. In addition, the training set for the design of the RBF networks consists of 500 (out of the 1491) signals, obtained from distinct speed and density environmental parameters.

A. Inversion performance using the exact design process P1

In this section, we study the performance of a RBF network designed using the exact process **P1**. The RBF network is constructed using the estimated $S\alpha S$ parameters of a subset containing $M \in \{50:50:300\}$ signals chosen from the training set. Besides, $N_1 = M$ since the RBF network is designed using the exact process. As mentioned before, the selection of the spread parameters $\{\sigma_i\}_{i=1,\ldots,N_1}$ is crucial for an improved performance of the inversion scheme. In the proposed method, we assume that all the radial basis neurons have equal spread parameter, that is, $\sigma_i = \sigma$, $\forall i$.

For the determination of σ, the heuristic described in Sec. II D is followed. In particular, in the case of the standard RBF network, an efficient design is ensured when $0 \leq \sigma$
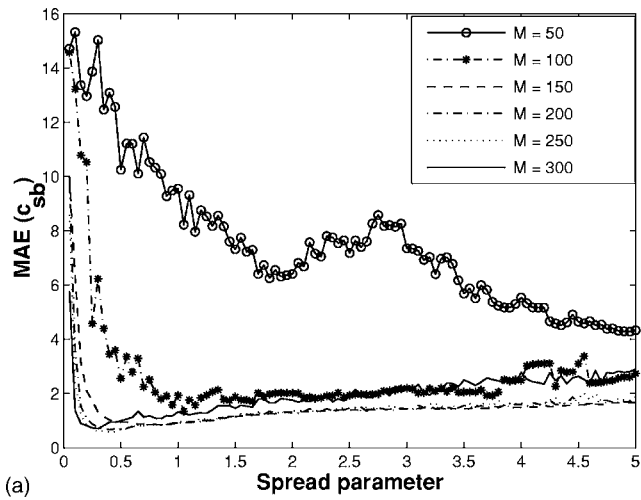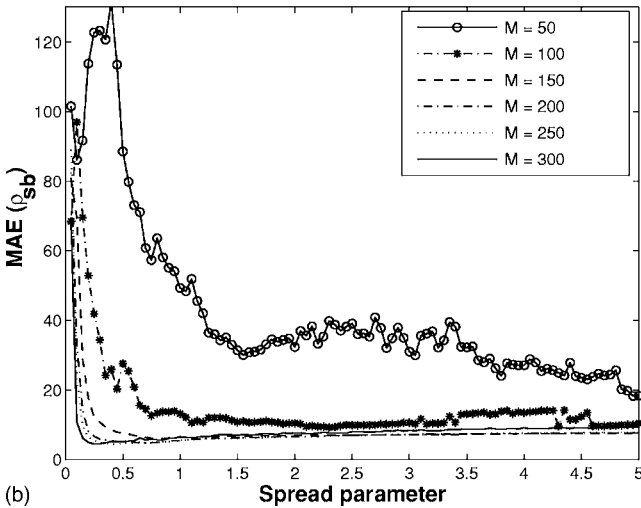
FIG. 4. Mean absolute error, given by the standard RBF network, as a function of the spread parameter σ and the number of training samples M for: (a) the sound speed c_{sb} and (b) the substrate density ρ_{sb}.

FIG. 5. Mean absolute error, given by the modified RBF network employing the standard KLD, as a function of the spread parameter σ and the number of training samples M for: (a) the sound speed c_{sb} and (b) the substrate density ρ_{sb}.

≤ 0.231, where 0 and 0.231 are the minimum and maximum Euclidean distances, respectively, between the $S\alpha S$ parameter vectors of all the signals in the database. Similarly, in the case of the modified RBF network, the value of σ should be selected such that $2.2204 \times 10^{-16} \leq \sigma \leq 252.59$, where 2.2204×10^{-16} and 252.59 are the minimum and maximum KLDs, respectively, given by Eq. (4), between the $S\alpha S$ parameter vectors of all the signals in the database. Finally, when the modified RBF network employs the symmetrized KLD as a distance function, one should choose $8.8818 \times 10^{-16} \leq \sigma \leq 259.49$, where 8.8818×10^{-16} and 259.49 are the minimum and maximum values of the symmetrized KLDs, respectively, given by Eq. (8).

The inversion performance is evaluated using several values of σ, as well as of the sharpness parameter r of the general transfer function $\phi(n) = e^{-n^r}$. In particular, in the case of the standard RBF network, the value of σ varies in the interval [0.05, 5], while in the case of the modified RBF it varies in the interval [0.001, 275]. In both cases, r belongs to the set $\{1, \ldots, 5\}$. For a given training size M and for fixed r and σ, we run 100 Monte Carlo iterations, where in each iteration a new RBF network is designed by randomly select-

ing a training subset of size M from the training set. In the subsequent figures and tables, the results have been obtained using a sharpness parameter $r=2$, otherwise, the value of r will be mentioned explicitly.
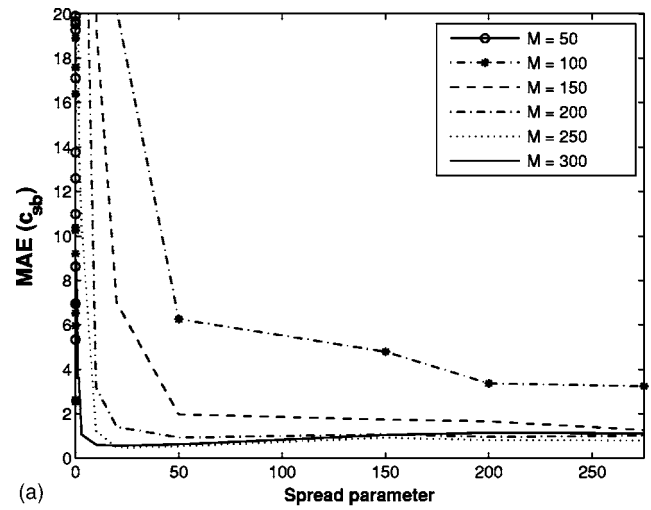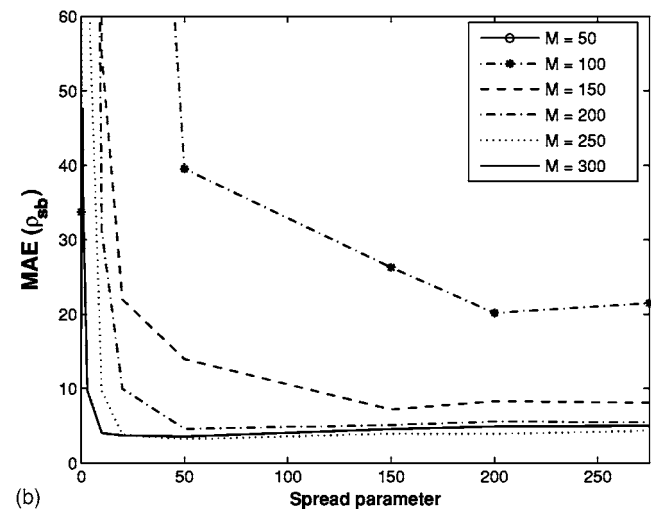
Figures 4(a) and 4(b) show the total mean absolute error (MAE) of the estimated sound speed c_{sb} and substrate density ρ_{sb} values, respectively, over the whole test set, given by the standard RBF network, as a function of the spread parameter σ and the number of training samples M. First, it can be seen that in both cases, for a relatively large number of training samples, the minimum MAE is achieved for a value of σ satisfying the corresponding inequality $0 \leq \sigma \leq 0.231$. The second observation is that, as expected, in this region the performance is improved, that is, the MAE decreases, as M increases, while out of this region this rule is not valid, especially for the estimation accuracy of c_{sb}.

Figures 5(a) and 5(b)[15] show the total MAE of the estimated c_{sb} and ρ_{sb} values, respectively, over the whole test set, given by the modified RBF network employing the standard KLD as a "distance" function, versus the spread parameter σ and for various training sample sizes M. It is clear that
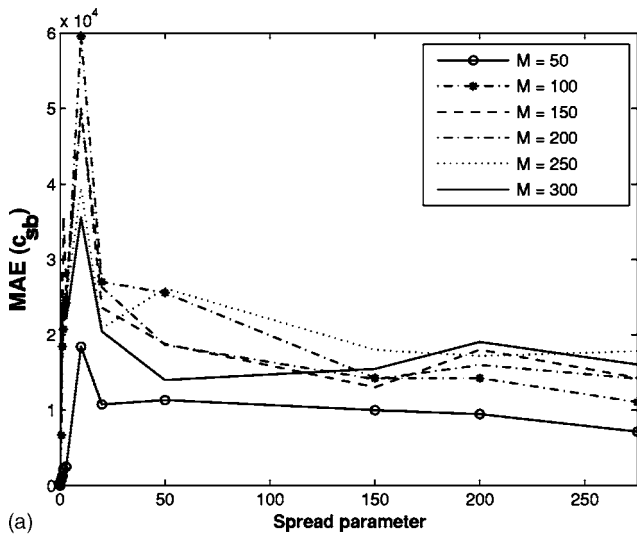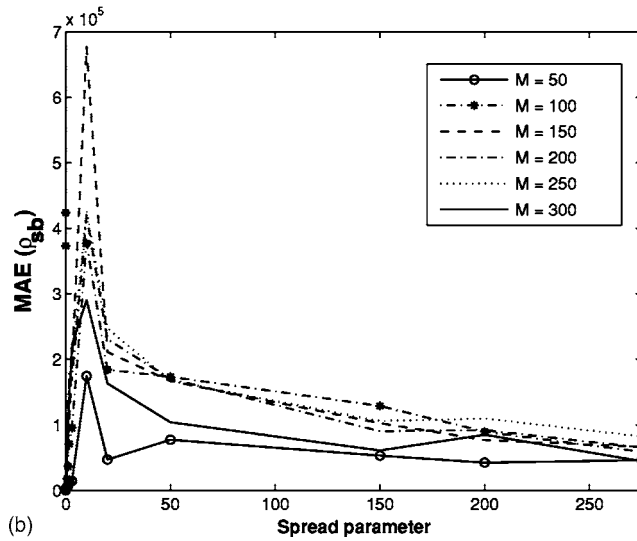
FIG. 6. Mean absolute error, given by the modified RBF network employing the symmetrized KLD, as a function of the spread parameter σ and the number of training samples M for: (a) the sound speed c_{sb}, and (b) the substrate density ρ_{sb}.

for both environmental parameters and a relatively large number of training samples, the minimum MAE is achieved for a value of σ satisfying the corresponding inequality

$2.2204 \times 10^{-16} \leqslant \sigma \leqslant 252.59$. Besides, it can be seen that the MAE decreases as M increases, as expected. Also note that in both cases, c_{sb} and ρ_{sb}, and for a fixed M, the performance of the modified RBF network stabilizes for $\sigma > 50$. Finally, Figs. 5(a) and 5(b) show that as M increases, the corresponding value of σ, which minimizes the MAE, decreases. This should be expected, since as the number of training samples increases, the input space can be covered using radial basis functions with a smaller spread.

Figures 6(a) and 6(b) show the total MAE of the estimated c_{sb} and ρ_{sb} values, respectively, over the whole test set, given by the modified RBF network employing the symmetrized KLD as a distance function. The first observation is that for both environmental parameters, the MAE is too large, compared with the performance of the other two RBF networks. Besides, the behavior of this network does not follow the rule that the MAE decreases as the number of training samples increases. In particular, there is a region on the x axis ([0,50]), where the MAE is minimized when the network is trained with a very small number of samples. On the other hand, similar to the case of the previous RBF network, which employs the standard KLD, the behavior of the MAE stabilizes for $\sigma > 50$, but the minimum MAE is still achieved for $M = 50$.

An explanation for this unexpected behavior of the modified RBF network based on the symmetrized KLD is shown in Fig. 7. Consider the simple case of a RBF network consisting of two hidden units with centers $\mathbf{c}_A = \mathbf{d}_A$ and $\mathbf{c}_B = \mathbf{d}_B$, respectively, where the vectors $\mathbf{d}_A$ and $\mathbf{d}_B$ contain the estimated $S\alpha S$ parameters of two signals S_A, S_B recorded in two very distinct sea environments. Also assume that the same vectors, $\mathbf{d}_A$ and $\mathbf{d}_B$, are given as inputs in the RBF network. As can be seen in Fig. 7, when the standard KLD is employed, the network outputs are different, $D(\mathbf{d}_A, \mathbf{d}_B) \neq D(\mathbf{d}_B, \mathbf{d}_A)$, since the standard KLD is not symmetric. On the other hand, if the symmetrized KLD is employed, we observe that the network gives the same output, although the input vectors correspond to two very distinct environments. Subsequently, the linear layer will result in an increased estimation error of the environmental parameters c_{sb}, ρ_{sb}.

Table II shows the average performance and the corresponding optimal design parameters for the three types of

FIG. 7. Comparison between the output of the hidden radial basis layer obtained by employing the standard and the symmetrized KLD.

Hidden Layer Output (H)		
	Standard KLD	Symmetrized KLD
(i)	$\mathrm{H} = D(\mathbf{d}_A, \mathbf{d}_A) + D(\mathbf{d}_A, \mathbf{d}_B)$ $= D(\mathbf{d}_A, \mathbf{d}_B)$	$\mathrm{H} = D_{sym}(\mathbf{d}_A, \mathbf{d}_A) + D_{sym}(\mathbf{d}_A, \mathbf{d}_B)$ $= 0 + \{D(\mathbf{d}_A, \mathbf{d}_B) + D(\mathbf{d}_B, \mathbf{d}_A)\}$
(ii)	$\mathrm{H} = D(\mathbf{d}_B, \mathbf{d}_B) + D(\mathbf{d}_B, \mathbf{d}_A)$ $= D(\mathbf{d}_B, \mathbf{d}_A)$	$\mathrm{H} = D_{sym}(\mathbf{d}_B, \mathbf{d}_A) + D_{sym}(\mathbf{d}_B, \mathbf{d}_B)$ $= \{D(\mathbf{d}_B, \mathbf{d}_A) + D(\mathbf{d}_A, \mathbf{d}_B)\} + 0$

TABLE II. Average error performance and optimal design parameters for the three types of RBF networks ($r=2$).

RBF type	M	c_{sb} (m/s)			ρ_{sb} (kg/m³)		
		σ	MAE	Std of error	σ	MAE	Std of error
Standard	250	0.4	0.5792	1.8933	0.45	4.4230	11.5974
Modified with $D(\cdot,\cdot)$	300	0.006	0.2481	0.7579	50	3.5687	2.5640
Modified with $D_{sym}(\cdot,\cdot)$	50	0.006	7.6462	3.4814	0.001	45.7241	25.4485

RBF networks. It is clear that the proposed modified RBF network, based on the standard KLD, outperforms the standard RBF network, based on the Euclidean norm, as well as the modified RBF network based on the symmetrized KLD. Besides, the increased estimation performance with respect to the substrate density becomes more evident, if we recall that the substrate density is one of the most difficult environmental parameters to be estimated by the majority of the previously developed inversion schemes. One should also note that distinct optimal σ parameters are necessary for estimating each environmental variable c_{sb}, or ρ_{sb}. Hence, for

optimal performance, two different single-output networks should be designed, one for each environmental parameter, using the corresponding optimal spread parameter. Of course, in this case there is a trade-off between the increased approximation accuracy and the computational complexity.

As mentioned before, we are also interested in studying the influence of the sharpness parameter (r) of a radial basis function, on the average performance of the proposed inversion scheme. Figure 8(a) shows the MAE of the estimated c_{sb} values using the standard RBF network, versus the spread σ,
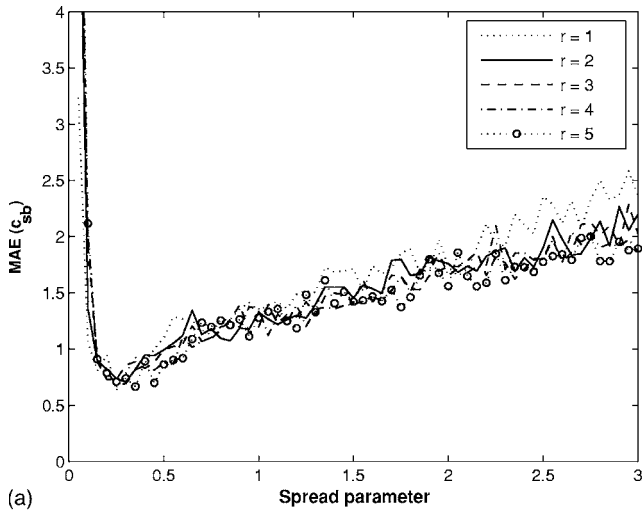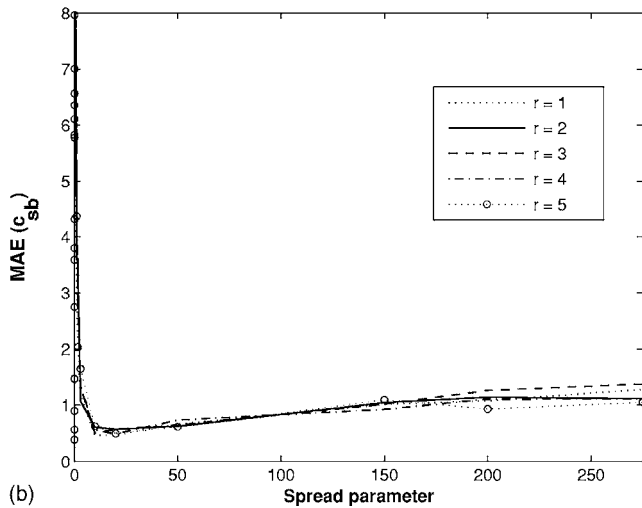
FIG. 8. Mean absolute error of the estimated c_{sb} values, as a function of the spread parameter σ and the sharpness parameter r, corresponding to: (a) the standard RBF network and (b) the modified RBF network employing the standard KLD.
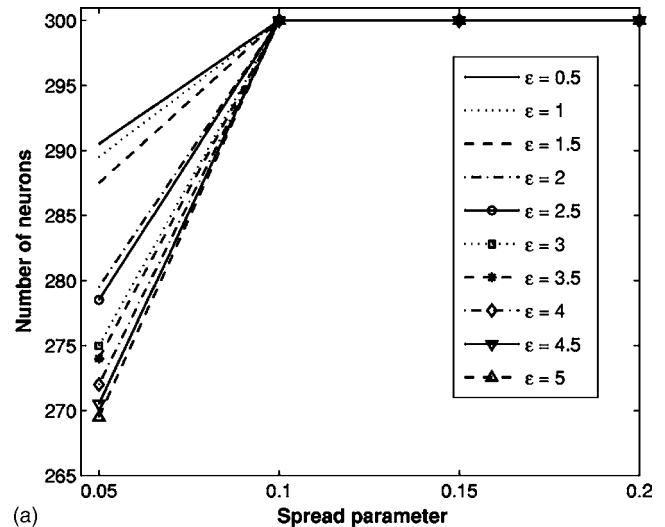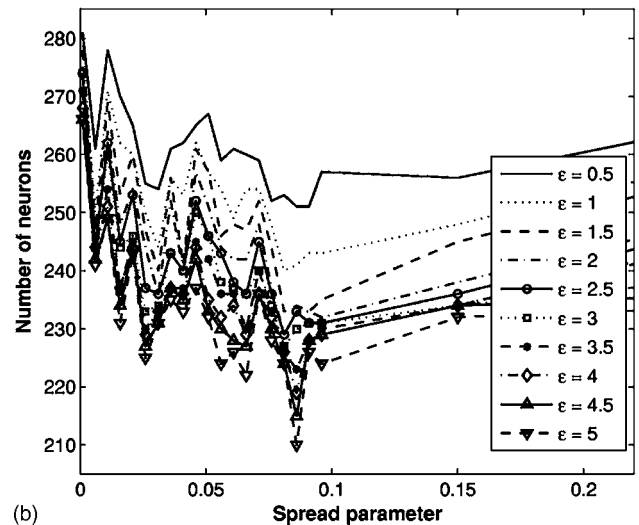
FIG. 9. Number of neurons as a function of the spread parameter σ and for several error thresholds ε, corresponding to: (a) the standard RBF network and (b) the modified RBF network based on the standard KLD.

Tzagkarakis *et al.*: Statistical geoacoustic inversion via neural networks

TABLE III. Comparison between the optimal number of neurons (N_1^{opt}) required to achieve a predetermined error threshold ε, for the standard and modified RBF networks ($r=2$).

	N_1^{opt}	
ε	Standard	Modified with $D(\cdot,\cdot)$
0.5	290	251
1	289	240
1.5	287	233
2	279	229
2.5	278	229
3	275	227
3.5	274	223
4	272	219
4.5	270	215
5	269	210

for various values of the sharpness parameter r. As can be seen, the performance is almost independent of the value of r. Figure 8(b) shows the average performance when the sound speed is estimated using the modified RBF network based on the standard KLD. It is clear that this network exhibits a higher degree of independence, with respect to the value of r, in comparison to the standard RBF network. Similar behavior was also found for both network types in the case of the substrate density estimation.

B. Inversion performance using the efficient design process P2

In this section, we study the performance of a network designed using the efficient process **P2**. The RBF network is constructed using the estimated $S\alpha S$ parameters of a subset containing $M=300$ signals chosen from the training set. Besides, let ε denote the error threshold, which we vary in the interval [0.5, 5]. In addition, the maximum number of neurons constituting the RBF network, if the error threshold is not achieved, is set equal to $M=300$. In the case of the standard RBF network, the value of σ varies in the interval [0.05, 3], while in the case of the modified RBF based on the standard KLD, it varies in the interval [0.001, 275]. In both cases, the value of the sharpness parameter is set to $r=2$. For a given error threshold ε and for a fixed σ, we run 100 Monte Carlo iterations, where in each iteration a new RBF network is designed by randomly selecting a training subset of size $M=300$ from the training set.

Figure 9(a) shows the number of neurons in the standard RBF network, as a function of σ and for several values of the error threshold ε. We can see that for each value of ε, the maximum number of neurons is required to meet the corresponding error threshold, when the spread parameter is greater than 0.1. On the other hand, the number of neurons required to achieve the error threshold decreases as the value of ε increases, for $\sigma < 0.1$. This behavior should be expected, since we need more neurons in order to achieve an increased resolution of the input space, for a given value of σ.

Figure 9(b) shows the number of neurons in the modified RBF network employing the standard KLD, as a function of σ and for several values of the error threshold ε. We
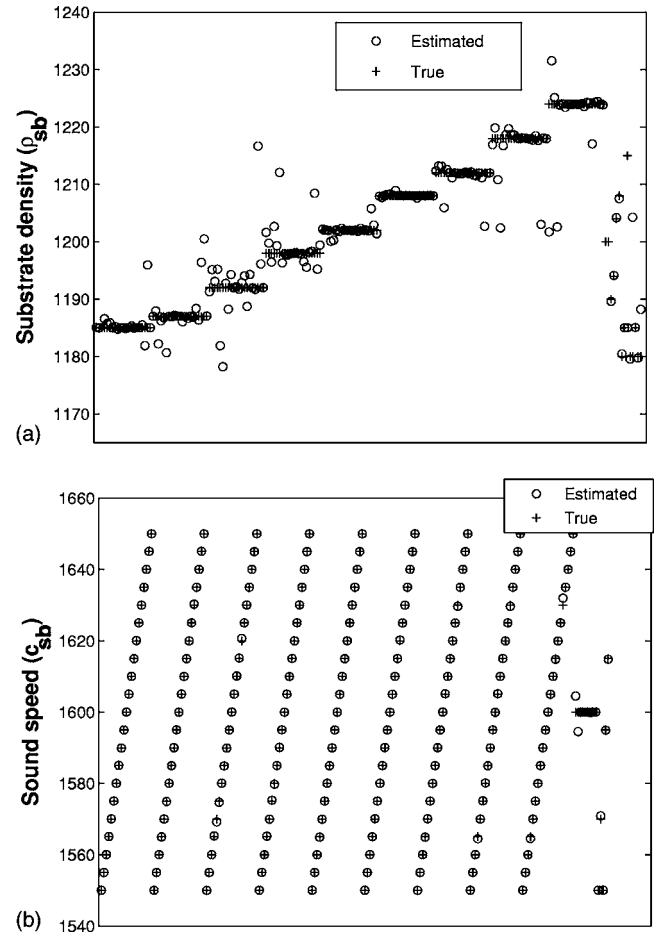
FIG. 10. True and estimated values of: (a) the substrate density and (b) the sound speed, obtained using the modified RBF network based on the standard KLD, with the optimal parameters of Table II.

can see that for each value of ε, the optimal number of neurons required to meet the corresponding error threshold does not follow the linear behavior as in the case of the standard RBF, but it strongly depends on the value of σ. However, the general rule that the number of neurons required to achieve the error threshold decreases as the value of ε increases, for a fixed σ, is still valid. Besides, by comparing Figs. 9(a) and 9(b), we can see that the proposed modified RBF network consists of less neurons than the standard RBF network, for the same error threshold ε. Table III SHOWS THE OPTIMAL number of neurons required to achieve the same error threshold, for both the standard and the modified RBF networks. It is clear that the proposed novel modified RBF network reduces significantly the number of hidden neurons, compared with its standard version. This reduction is important, since the number of hidden neurons affects the computational complexity of a neural network.

Finally, Figs. 10(a) and 10(b) show the true and estimated values for the substrate density and the sound speed, respectively, for a set of synthetic signals of our database (the horizontal axis is simply the signal index and, thus, omitted).[16] The estimated values were obtained using the proposed modified RBF network based on the standard KLD, which is designed using the optimal parameters of Table II.

As was expected, the estimation accuracy of the sound speed is higher than the accuracy of the estimation of the substrate density.

IV. CONCLUSIONS

In this paper, an inversion scheme for acoustic signals recorded in shallow water was presented and evaluated, based on a $S\alpha S$ modeling of the coefficients of the 1D wavelet decomposition, followed by use of a novel RBF neural network. In particular, we demonstrated that the parameters of the $S\alpha S$ distributions constitute an effective set of features that can be employed for building a RBF neural network, which exploits the KLD between $S\alpha S$ distributions. This modified RBF network recovers efficiently the unknown environmental parameters of the recorded signal, achieving a decreased average error compared with the standard RBF network. Besides, our proposed RBF network requires less hidden neurons to achieve the same error threshold with its standard version. Our future work consists of testing the proposed scheme in real shallow-water environments, when the received signal is contaminated with noise. We also plan to study other neural network structures in order to achieve an even better performance of the inversion scheme.

[1] *Inverse Problems in Underwater Acoustics*, edited by M. I. Taroudakis and G. N. Makrakis (Springer, Berlin, 2001).
[2] *Full Field Inversion Methods in Ocean and Seismic Acoustics*, edited by O. Diachok, A. Caiti, P. Gerstoft, and H. Schmidt (Kluwer Academic, Dordrecht, 1995).
[3] "Benchmarking geoacoustic inversion methods," J. Comput. Acoust. (Special edition, K. Chapman and A. Tolstoy, eds.) **6**, 1–289 (1998).
[4] N. R. Chapman, S. Chin-Bing, D. King, and R. B. Evans, "Benchmarking geoacoustic inversion methods for range-dependent waveguides," IEEE J. Ocean. Eng. **28**, 320–330 (2003).
[5] M. Taroudakis, G. Tzagkarakis, and P. Tsakalides, "Classification of shallow-water acoustic signals via alpha-Stable modeling of the 1-D wavelet coefficients," J. Acoust. Soc. Am. **119**, 1396–1405 (2006).
[6] M. Taroudakis, G. Tzagkarakis, and P. Tsakalides, "Characterization of an under-water acoustic signal using the statistics of the wavelet subband coefficients," in *Theoretical and Computational Acoustics*, edited by A. Tolstoy, E.-C. Chang, and Y.-C. Teng (World Scientific, Singapore, 2006), pp. 167–174.
[7] Y. Stephan, X. Demoulin, and O. Sarzeaud, "Neural direct approaches for geoacoustic inversion," J. Comput. Acoust. **6**, 151–166 (1998).
[8] J. Benson, N. R. Chapman, and A. Antoniou, "Geoacoustic model inversion using artificial neural networks," Inverse Probl. **16**, 1627–1639 (2000).
[9] S. Mallat, *A Wavelet Tour of Signal Processing* (Academic, New York, 1998).
[10] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," IEEE Trans. Pattern Anal. Mach. Intell. **11**, 674–692 (1989).
[11] J. P. Nolan, "Parameterizations and modes of stable distributions," Stat. Probab. Lett. **38**, 187–995 (1998).
[12] J. P. Nolan, "Numerical calculation of stable densities and distribution functions," Comput. Stat. Data Anal. **13**, 759–774 (1997).
[13] S. Kullback, *Information Theory and Statistics* (Dover, New York, 1997).
[14] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. (Prentice Hall, Englewood Cliffs, NJ, 1998).
[15] Notice that in both Figs. 5(a) and 5(b) the curve corresponding to M =50 is not completely shown, since the y axis was scaled for a better visualization.
[16] Notice that the staircase-like view of Fig. 10(a) and the piecewise linear view of Fig. 10(b) is simply due to the way we grouped the signals in the data set. For convenience in the visual interpretation, we grouped the signals as follows: the first group contains signals with c_{sb} $\in [1550, 1650]$ m/s and the same density ρ_{sb}=1185 kg/m^3, the second group contains signals with $c_{sb} \in [1550, 1650]$ m/s and the same density ρ_{sb}=1187 kg/m^3, and so on.

Sperm whale three-dimensional track, swim orientation, beam pattern, and click levels observed on bottom-mounted hydrophones[a]

Eva-Marie Nosal[b] and L. Neil Frazer

School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, Hawaii 96822

In an earlier paper [Nosal and Frazer Appl. Acoust. **61**, 1187–1201 (2006)], a sperm whale was tracked in three-dimensions using direct and surface-reflected time differences (DRTD) of clicks recorded on five bottom-mounted hydrophones, a passive method that is robust to timing errors between hydrophones. This paper refines the DRTD method and combines it with a time of (direct) arrival method to improve the accuracy of the track. The position and origin time of each click having been estimated, pitch and yaw are then obtained by assuming the main axis of the whale is tangent to the track. Roll is then found by applying the bent horn model of sperm whale phonation, in which each click is composed of two pulses, p0 and p1, that exit the whale at different points. With instantaneous pitch, roll, and yaw estimated from time differences, amplitudes are then used to estimate the beam patterns of the p0 and p1 pulses. The resulting beam patterns independently confirm those obtained by Zimmer *et al.* [J. Acoust. Soc. Am. **117**, 1473–1485 (2005); **118**, 3337–3345 (2005)] with a very different experimental setup. A method for estimating relative click levels is presented and used to find that click levels decrease toward the end of a click series, prior to the "creak" associated with prey capture.
© *2007 Acoustical Society of America.* [DOI: 10.1121/1.2775423]

I. INTRODUCTION

The main purpose of this paper is to demonstrate the progress and use of passive acoustic methods for studying marine mammals in the wild, especially odontocetes. In a recent paper (Nosal and Frazer, 2006), we studied the improvement in ray-based tracking that occurs when a realistic sound speed profile is used instead of an assumed isospeed profile. We tracked a sperm whale using the difference between direct and surface-reflected click arrival times (DRTD), a method that is robust to time-origin errors on different hydrophones. Here we refine the DRTD method and combine it with a time of (direct) arrival (TOA) method to get a combined method that is more accurate than either method separately. The time and position estimates are precise enough that we can approximate swim velocity and orientation at each click, which we then use to estimate click beam patterns and levels.

This paper focuses on clicks with regular interclick intervals of 0.45–2 s, called "usual" clicks by Whitehead and Weilgart (1990), emitted by sperm whales while foraging at depth. A typical foraging dive lasts about 45 min, and begins with a steep, steady descent to a depth of several hundred meters (Watkins *et al.*, 2002; Watwood *et al.*, 2006), followed by a period of searching for prey and then a steep,

steady ascent. Series of regular clicks emitted at foraging depth are often terminated by a "creak" of clicks with high repetition rate followed by several seconds of silence (Gordon, 1987; Mullins *et al.*, 1988; Goold and Jones, 1995). The regular clicks are likely used for echolocation (Norris and Harvey, 1972; Whitehead and Weilgart, 1991; Goold and Jones, 1995; Møhl *et al.*, 2000; Jaquet *et al.*, 2001; Madsen *et al.*, 2002a; Whitehead, 2003; Møhl *et al.*, 2003) while the creaks mark the terminal phase of prey capture (Miller *et al.*, 2004). Regular clicks are short in duration, broadband (100 Hz to over 20 kHz), and have a powerful forward directed beam (Møhl *et al.*, 2000; Madsen *et al.*, 2002a, b). Regular clicks are often heard on hydrophones several kilometers from the vocalizing animal. Being of short duration, direct and reflected arrivals can often be distinguished, making clicks ideal for passive localization.

We track a single sperm whale from its regular clicks for 23 min using recordings on five bottom-mounted hydrophones. The data were recorded at the Atlantic Undersea Test and Evaluation Center (AUTEC) located in the Tongue of the Ocean (off Andros Island, Bahamas). They were provided by the Naval Undersea Warfare Center for the second International Workshop of Detection and Localization of Marine Mammals using Passive Acoustics. The sampling rate was 48 kHz and the hydrophone positions are listed in Table I. Some further details can be found in Adam *et al.* (2006), but unfortunately, the anti-alias filter, frequency response, sensitivity, and directionality of the sensors were not available. Accordingly, our results are limited by the assumption of an omnidirectional and flat frequency response, and absolute sound pressure levels cannot be found.

TABLE I. Hydrophone positions provided by NUWC.

Hydrophone	x position (m)	y position (m)	Depth (m)
G	10 658.04	−14 953.63	1530.55
H	12 788.99	−11 897.12	1556.14
I	14 318.86	−16 189.18	1553.58
J	8 672.59	−18 064.35	1361.93
K	12 007.50	−19 238.87	1522.54

The data set used here is the same data set that we used to develop the DRTD method (Nosal and Frazer, 2006). The track of the sperm whale in this data set has also been obtained using time of arrival differences between pairs of receivers (Giraudet and Glotin, 2006; Morrissey et al., 2006; White et al., 2006). The improved accuracy of the combined DRTD/TOA method used in this paper allows us to estimate the velocity of the sperm whale from position and time differences between successive clicks. The pitch and yaw of the whale then follow from the assumption that the main axis of the whale is parallel to its velocity vector.

In order to find roll, we apply the bent-horn model of sperm whale phonation, which was proposed to explain the multipulse structure of sperm whale clicks (Norris and Harvey, 1972; Møhl, 2001) and is supported by recent studies (Madsen et al., 2003; Møhl et al., 2003; Zimmer et al., 2005b). In the bent-horn model, a single sound is generated at the phonic lips (Fig. 1). Some energy leaks directly into the water as a p0 pulse. Most of the energy transmits back through the spermaceti organ, reflects off the frontal sac in front of the skull, transmits forward into the "junk," and exits into the water as the p1 pulse. Since the p1 pulse follows a longer path than p0, it arrives later, giving the click a multi-pulse structure. Other click components (resulting from further reflections in the head and other exit points) are also present (Møhl, 2001), but they are not required or considered here. Since the measured delay between the p0 and p1 pulses depends on the orientation of the whale (Zimmer et al., 2005a; Laplanche et al., 2006), it can be used to recover roll.

With position, velocity, pitch, roll and yaw obtained solely from travel time differences, we then use relative amplitudes to estimate the beam patterns and directivity indices of the p0 and p1 pulses that comprise a click. Our results agree with previous studies (Møhl et al., 2003; Zimmer et al., 2005b) which found that the p1 pulse has a narrowly focused, forward-direct beam, that the p0 pulse is slightly
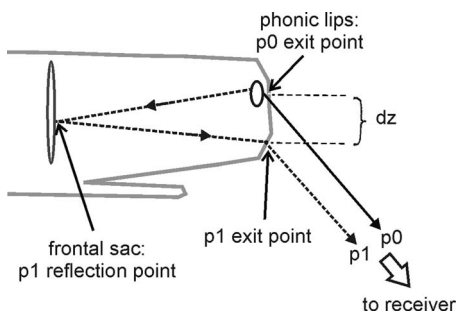
FIG. 1. Diagram illustrating the paths taken by the p0 and p1 pulses according to the modified bent-horn model of sound production in sperm whales.

weaker and more broadly backward-directed, and that a low-frequency, nearly omnidirectional component is characteristic of all clicks. Finally, we correct click amplitudes for beam pattern and propagation loss to estimate relative click levels within each click sequence, finding that click source levels decrease toward the end of a click series.

II. DATA PROCESSING

A. Click detection, classification, and association

The beginnings and ends of the clicks were detected using an automated transient detector (Page, 1954; Wald, 1947; Abraham, 2000; Zimmer et al., 2003, 2005b). To reduce noise, each time series was filtered using a second-order, high-pass Butterworth filter with a 300 Hz low cut. The envelope of each filtered time series was calculated as the magnitude of the corresponding analytic signal, where the analytic signal has real and imaginary parts consisting of the original time series and its Hilbert transform, respectively (e.g., Bracewell, 2000). Given the instantaneous signal amplitude (envelope) e_n, a test variable V_n was calculated as

$$V_n = \frac{e_n^2}{N_n},\tag{1}$$

where N_n is the noise estimate. For the first noise estimate, N_1 is taken as the mean square envelope value (over all samples for that hydrophone). Detection decisions and updates for subsequent noise estimates are made according to the value of V_n in the following algorithm

$$\text{If}\begin{cases}V_n > T_0\\V_n < T_x\\\text{other}\end{cases}\text{then}\begin{cases}\text{decide detection}\\\text{set }T_x = T_1\\\text{set }N_{n+1} = N_n\\[6pt]\text{decide no detection}\\\text{set }T_x = T_2\\\text{set }N_{n+1} = (1-\alpha)N_n + \alpha e_n^2\\[6pt]\text{no decision}\\\text{keep current }T_x\\\text{set }N_{n+1} = N_n\end{cases}\tag{2}$$

where T_0, T_1, and T_2 are the thresholds for decision of detection, end of detection, and noise, respectively; $T_x \in \{T_1, T_2\}$ is the current threshold; and α is the exponential weighting on the power estimate when no signal is detected. For the first sample, T_x is set equal to T_1. Threshold and weighting values that performed well were $T_0 = 25(13.98$ dB$)$, $T_1 = 9(9.54$ dB$)$, $T_2 = 4(6.02$ dB$)$, and $\alpha = 1/100$.

At each time step, this algorithm decides if there is a signal present (detection) or not (no detection). The algorithm operates in two modes: signal and noise. In the signal mode, signal present is decided while the value of the test variable is greater than the detection threshold ($V_n > T_0$). No decision is made if the value of the test variable is less than the threshold for detection but greater than the threshold for

the end of detection ($T_1 < V_n < T_0$). Signal not present is decided once the value of the test variable drops below the end of detection threshold ($V_n < T_1$). Here, the algorithm switches to noise mode. In this mode, signal not present is decided while the test variable remains below the noise threshold ($V_n < T_2$) and the noise variable is updated at each time step according to an exponential weighting (more weight toward recent values). No decision is made if the value of the test variable is less than the threshold for detection but greater than the threshold for noise ($T_2 < V_n < T_0$). Signal present is decided once the value of the test variable jumps above the detection threshold ($V_n > T_0$). Here, the algorithm switches to signal mode.

Each click resulted in a direct arrival, usually followed by a lower-amplitude surface reflection. Direct-reflected pairs were classified according to the following criteria: (1) the amplitude of the direct arrival varies slowly; (2) the interclick interval between successive direct arrivals varies slowly; (3) the reflected arrival has lower amplitude than the direct arrival; and (4) the time between the direct and reflected arrival (DRTD) is similar to that of the preceding direct-reflected pair.

Each click detected on two or more receivers was numbered sequentially. Clicks on different receivers were associated by comparing intervals between clicks, which should be nearly identical on all receivers. To eliminate incorrect associations due to click time measurement error, this comparison included intervals between all clicks in a series, not only those immediately preceding or following a given click. In total, 1324 clicks were numbered, with 1102, 913, 868, 1163, and 1035 clicks detected on receivers 1, 2, 3, 4, and 5, respectively. The number of clicks detected on a total of two, three, four, and five receivers was 137, 324, 480, and 383. Only clicks recorded on three or more hydrophones (total of 1187 clicks) were used for localization.

B. Level and pulse delay measurement

The maximum of the envelope was used to estimate the received peak pressures of each direct click. Using frequency and time windows, peak pressures of the p0, p1, and low-frequency (LF) components were also obtained. Following Zimmer *et al.* (2005b), the p0 and p1 pulses were defined to fall in time windows from, respectively, −2 to 3 ms and 3 to 10 ms, relative to detection of the start of the click. They were both defined to fall in a frequency window of 3–22 kHz. The identified p0 and p1 times and amplitudes corresponded to maxima of the envelope of the filtered signal. The LF component was defined with a time window of −2 to 10 ms and a frequency window of 300 Hz to 3 kHz. The delay between the p0 and p1 pulses, τ, was estimated by subtracting p0 arrival time from p1 arrival time.

III. LOCALIZATION

A. Methods

For each receiver, the acoustic propagation model BELL-HOP (Porter, 2005) was used to create a lookup table of TOAs, DRTDs, takeoff beam angles, and transmission losses for a list of candidate source ranges and depths. The historic

depth-dependent sound speed profile for the area (24° 45′ N, 77° 45′ W) in March was taken from the Generalized Digital Environment Model (GDEM) and is the same as the profile used in Nosal and Frazer (2006). The depth list varied from 5 to 1550 m with 5 m increments, and the range list varied from 5 m to 20 km with 5 m increments. Since arrival times varied smoothly for the depths and ranges of interest, all required TOAs and DRTDs were interpolated from the values in the lookup table using cubic splines.

To determine the time and position of each click, we first created a four-dimensional grid of candidate source points (one dimension for time and three for position). Errors in DRTD and TOA were assumed to be normally distributed. Ideally, DRTD and TOA should be regarded as functions, not just of source position, but also of sound speed and receiver positions, and likelihood surfaces should be maximized over this much larger parameter space. However, to reduce computational requirements, we incorporated the uncertainties in sound speed profile and receiver positions into the standard deviations for DRTD and TOA in a "worst case" manner. Standard deviations σ_{DRTD} and σ_{TOA} were calculated as

$$\sigma_{\text{DRTD}} = \sqrt{2\sigma_{\text{meas}}^2 + \sigma_{\text{rp}}^2 + \sigma_{\overline{\text{DRTD}}}^2}, \tag{3}$$

$$\sigma_{\text{TOA}} = \sqrt{\sigma_{\text{meas}}^2 + \sigma_{\text{rp}}^2 + \sigma_{\overline{\text{TOA}}}^2}, \tag{4}$$

where σ_{meas} is the standard deviation (s.d.) in the measured click times, σ_{rp} is the s.d. due to uncertainty in receiver position, and $\sigma_{\overline{\text{DRTD}}}$ and $\sigma_{\overline{\text{TOA}}}$ are the s.d. (due to sound speed uncertainty) in modeled DRTDs and TOAs. We used $\sigma_{\text{meas}} = 5$ ms based on the widths of the clicks (about 10 ms) and $\sigma_{\text{rp}} = 2$ ms corresponding approximately to a best-guess receiver position uncertainty of 3 m (actual position uncertainty is unknown). To determine $\sigma_{\overline{\text{DRTD}}}$ and $\sigma_{\overline{\text{TOA}}}$, the DRTD and TOA lookup tables were recalculated for all 12 months using historic sound speed profiles (also from the GDEM). This gave 12 possible TOAs and DRTDs for each range and depth. The difference between the minimum and the maximum of these 12 values approximates the width of the uncertainty curves. The maximum such width over all ranges and depths ("worst-case") was taken as 2 s.d., giving $\sigma_{\overline{\text{DRTD}}} = \sigma_{\overline{\text{TOA}}} = 3$ ms. Using the maximum width simplifies the calculations, by allowing 1 s.d. to be used for all candidate points, and it overestimates final errors.

For candidate whale position **s** and click time t, the DRTD and TOA likelihood functions were computed as

$$L_{\text{DRTD}}(\mathbf{s}) = \frac{1}{(2\pi\sigma_{\text{DRTD}}^2)^{N/2}}$$
$$\times \exp\left[\frac{-1}{2\sigma_{\text{DRTD}}^2}\sum_j (\overline{\text{DRTD}}_j(\mathbf{s}) - \text{DRTD}_j)^2\right], \tag{5}$$

$$L_{\mathrm{TOA}}(\mathbf{s},t) = \frac{1}{(2\pi\sigma_{\mathrm{TOA}}^2)^{N/2}}$$
$$\times \exp\left[\frac{-1}{2\sigma_{\mathrm{TOA}}^2}\sum_j (\overline{\mathrm{TOA}}_j(\mathbf{s},t) - \mathrm{TOA}_j)^2\right], \tag{6}$$

where the sums are over all receivers that heard the click, N is the number of receivers that heard the click, $\underline{\mathrm{DRTD}}_j$ and $\underline{\mathrm{TOA}}_j$ are the measured values on receiver j, and $\overline{\mathrm{DRTD}}_j$ and $\overline{\mathrm{TOA}}_j$ are the modeled values. The total likelihood value is the product of these:

$$L(\mathbf{s},t) = L_{\mathrm{DRTD}}(\mathbf{s})L_{\mathrm{TOA}}(\mathbf{s},t). \tag{7}$$

The point $(\mathbf{s},t)$ that maximizes L is the estimated source position and time. An advantage of distinct likelihood surfaces is that they can be examined separately as a diagnostic, since persistent differences between locations from the two methods are an indication that hydrophone time origins may be different (degrading TOA), or that the sound speed profile in the upper part of the water column is inaccurate (degrading DRTD).

For computational efficiency, two passes were made. The first pass was coarsely sampled in space (10 m grid spacing) and time (10 ms time spacing). For the first click, the spatial search volume covered the full water column in depth and extended 5 km past the boundary defined by the receivers. Time was searched from 0 to 20 s. For the other clicks the boundary of the search volume was based on the time, Δt, between the current click and the preceding localized click. This was estimated from the measured time between these two clicks on a single phone that heard both clicks. The search volume was centered on the position estimate of the previous click, and bounded in all three directions by double the maximum possible swim distance in Δt, i.e., $8\Delta t$ (assuming a swim speed of at most 4 m/s). Time was searched from the previously localized click until $2\Delta t$ after it.

The second pass refined the position and time estimate from the first pass by searching a smaller, more finely sampled, volume centered on the position and time found in the first pass. The search volume for this pass was sampled at intervals of 1 m in space and 1 ms in time. It was bounded in space by the coarsely determined source location, plus or minus 20 m in all directions, and in time by 200 ms before and after the coarsely determined click time.

B. Error estimates

The literature on bioacoustic localization arrays contains various approaches to quantify error (e.g., Whalberg et al., 2001; Spiesberger and Wahlberg, 2002). Since the complete likelihood surfaces were already calculated in the above-noted localization step, we applied a somewhat different approach, using the likelihood surfaces to give error estimates.

95% confidence intervals (CIs) were estimated from conditional likelihood functions (CLFs) by identifying the interval containing 2.5%–97.5% of the cumulative likelihood for the parameter of interest. For example, to find the confi-
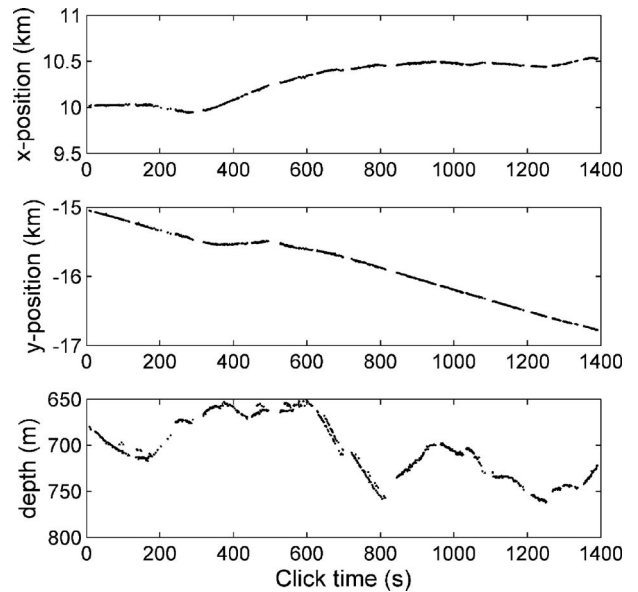
FIG. 2. Estimated track (not smoothed) of the sperm whale. Positions of clicks detected on three or more receivers are plotted as dots against the time of the click.

dence interval in the x position for a single click, all other parameters (y position, z position, and click time) were fixed to their values (y_0, z_0, t_0) at the estimated source position and time. The corresponding CLF, $L(x_j | y_0, z_0, t_0)$, was calculated according to Eq. (7) for a list of possible x positions, x_j. The cumulative CLF was then calculated as

$$C(x) \approx \frac{\sum\limits_{x_j < x} L(x_j)}{\sum\limits_{x_j} L(x_j)}. \tag{8}$$

The denominator normalizes the distribution, and the equality is approximate because of the discrete sampling of x. The list of x positions ranged from $x_0 - 1$ km to $x_0 + 1$ km (since the CLF was very close to 0 at 1 km away from x_0), with 1 m resolution. Then the 95% CI is $[x_{2.5\%}, x_{97.5\%}]$ where $x_{2.5\%}$ and $x_{97.5\%}$ are such that $C(x_{2.5\%}) = 0.025$ and $C(x_{97.5\%}) = 0.975$. CIs for y position, z position, and time were computed similarly.

C. Results

Figure 2 shows the resulting x, y, and z positions obtained for clicks heard on three or more receivers. Figure 3 shows the positions in the x-y plane. The click time list ranged from $t_0 - 1$ s to $t_0 + 1$ s with 0.1 ms resolution. The resulting CI half-widths for position are shown in Fig. 4. The half-widths for time were less than 4.5, 5, and 5.8 ms for clicks heard on five, four, and three receivers, respectively.

IV. SWIM ORIENTATION

A. Conventions

This section outlines the orientation conventions used in the following. Two reference frames are required: the earth frame and the whale frame, notated as unprimed and primed, respectively. In the earth frame, positive x, y, and z are di-
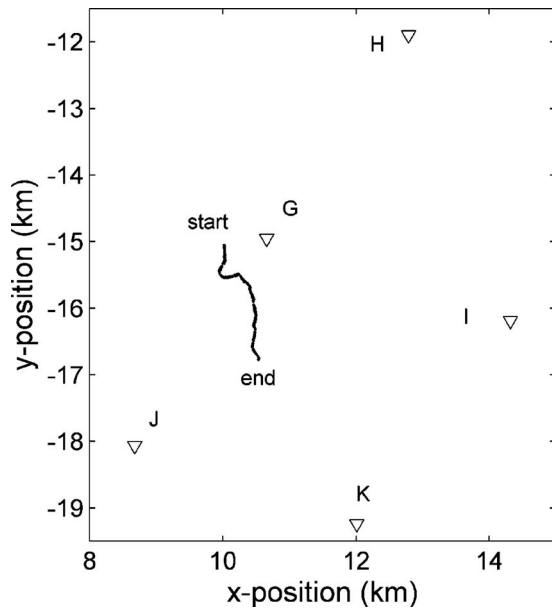
FIG. 3. Estimated two-dimensional track (not smoothed) of the sperm whale. Positions are plotted as dots to form the track, as in Fig. 2, and triangles indicate receiver positions.

rected east, north, and upward, respectively. In the whale frame, positive x', y', and z' point forward (rostrally along the whale's long axis), left, and dorsally, respectively. The two frames coincide when the whale is traveling due east, horizontal and upright.

Three angles are required to transform between the whale and earth frames: yaw (θ_z), pitch (θ_y), and roll (θ_x), which are rotations about the z, y, and x axes, respectively. For yaw and roll, positive values correspond to a coordinate system rotation in a clockwise direction when looking away from the origin along the axis of rotation. For consistency with conventions used by Johnson and Tyack (2003), and so

FIG. 4. Half-widths of 95% confidence intervals for position. Red, green, and blue indicate clicks detected on five, four, and three receivers, respectively.

that a positive pitch corresponds to a nose-upward orientation, positive pitch corresponds to a counter-clockwise rotation when looking away from the origin along the y axis. Note that this convention differs from standard Euler and pitch-roll-heading convention (Goldstein, 1980). Thus in our convention a whale with zero yaw, pitch, and roll is swimming eastward, horizontally, and upright. From this $\theta_x = \theta_y = \theta_z = 0$ orientation, the whale turns left to increase yaw, toward the surface to increase pitch, and clockwise to increase roll. To make them unique, θ_x, θ_y, and θ_z are constrained to the intervals $(-180°, 180°]$, $[-90°, 90°]$, $(-180°, 180°]$, respectively.

A vector in the earth (unprimed) frame is expressed in whale (primed) frame coordinates via three matrices that commute only in the limit of very small angles (so the order of multiplication is important):

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R_x(\theta_x) R_y(\theta_y) R_z(\theta_z) \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{9}$$

with

$$R_x(\theta_x) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & \sin\theta_x \\ 0 & -\sin\theta_x & \cos\theta_x \end{bmatrix}, \tag{10}$$

$$R_y(\theta_y) = \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{bmatrix}, \tag{11}$$

$$R_z(\theta_z) = \begin{bmatrix} \cos\theta_z & \sin\theta_z & 0 \\ -\sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{12}$$

B. Pitch and yaw

The first step in recovering swim attitude is to approximate the velocity of the whale at each click. To do this, a vector-valued position function $\mathbf{f}(t) = (f_x(t), f_y(t), f_z(t))$ was fit to the calculated click positions and times by minimizing a weighted sum of squared position error and acceleration:

$$E(\mathbf{f}) = a \sum_{j=1}^{N_c} |\mathbf{s}_j - \mathbf{f}(t_j)|^2 + (1-a) \int_{t_1}^{t_{N_c}} \left| \frac{d^2\mathbf{f}}{dt^2} \right|^2 dt, \tag{13}$$

where the sum is over all localized clicks j, N_c is the total number of localized clicks, $\mathbf{s}_j$ and t_j are the estimated position and time of click j, and $a = 0.7$ is a smoothing parameter. Velocity in the earth frame, $\mathbf{v}(t) = (v_x(t), v_y(t), v_z(t))$, is found by taking the first derivative of $\mathbf{f}$.

To recover pitch and yaw, we assume that the whale's main axis is parallel to its velocity. This assumption neglects the effects of current and the ability of the whale to move laterally or vertically, as well as any scanning movements of the head, so the goodness of the approximation increases with the forward speed of the whale. Pitch and yaw can then be computed as

$$\theta_y = \sin^{-1}(v_z/|\mathbf{v}|),\qquad(14)$$

$$\theta_z = \tan^{-1}(v_y/v_x),\qquad(15)$$

where $\tan^{-1}$ is the four-quadrant, inverse tangent.

C. Roll

Once position, pitch, and yaw are known, roll is estimated from the delay τ between the p0 and p1 pulses (Fig. 1). To do this, we build on methods introduced by Zimmer et al. (2005a) and Laplanche et al. (2006). The modeled delay is split into two components: $\tau = t_c + t_{\Delta p}$. The constant component, t_c, is the time required for sound to travel from the phonic lips to the frontal sac, where it reflects, and thence forward to the p1 exit point. It is assumed fixed for a given animal. The second component, $t_{\Delta p}$, is the difference between the travel time from the p1 exit point to the receiver and the travel time from the p0 exit point to the receiver. It depends on the exit point of the p1 pulse relative to the phonic lips—we assume this exit point is fixed for a given animal—and on whale position, receiver position, and roll. The exit point of the p1 pulse is located at the junk and is directly ventral to the phonic lips (Madsen, 2002; Zimmer et al., 2005a). Hence, if we take the phonic lips to be at point (0,0,0) in the whale frame, the p1 exit point can be approximated as $(0,0,-dz)$, with dz fixed for the individual whale (Fig. 1).

Given the position of the whale, $\mathbf{s}$, and the receiver, $\mathbf{r}$, in the earth frame, as well as the pitch and yaw angles previously determined, τ can be modeled (for various t_c, dz, and roll angles) as follows. For each click and receiver, we find the takeoff direction of the ray that connects the receiver and localized source positions. For a constant sound-speed profile, this direction in the earth frame would simply be $\mathbf{r}-\mathbf{s}$ with source position $\mathbf{s}$ and receiver position $\mathbf{r}$. To find this direction for our depth-dependent sound speed profile, the list of takeoff angles from BELLHOP is interpolated to get the vertical angle (or elevation), ϕ, of the ray in the earth frame. The ray direction vector in the earth frame is then

$$\mathbf{b} = [r_x - s_x,\ r_y - s_y,\ \tan(\phi)\sqrt{(r_x - s_x)^2 + (r_y - s_y)^2}]^T.\quad(16)$$

This ray direction is transformed into the whale frame direction vector, $\mathbf{b}'$, by applying Eq. (9) with the calculated yaw, pitch, and candidate roll angle for the current click. The azimuth $\alpha' \in [-180°, 180°]$ and elevation $\phi' \in [-90°, 90°]$ of the ray in whale coordinates are then calculated as

$$\alpha' = \tan^{-1}\left(\frac{b'_y}{b'_x}\right),\quad \phi' = \tan^{-1}\left(\frac{b'_z}{\sqrt{(b'_x)^2 + (b'_y)^2}}\right),\qquad(17)$$

where $\tan^{-1}$ is the four-quadrant inverse tangent. Here, positive/negative azimuth corresponds to a leftward/rightward directed beam, and positive/negative elevation corresponds to an upward/downward directed beam. Elevation and azimuth of 0° correspond to a beam directed along the whale's main axis, x'.

Since the distance from the whale to the receiver is much greater than dz, the vertical takeoff angles of the p0
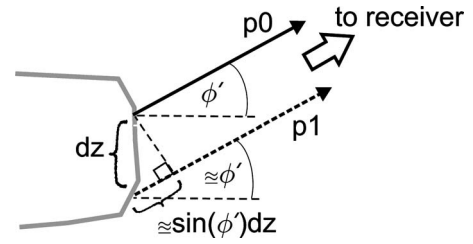
FIG. 5. Angle approximation used to estimate the delay between p0 and p1.

pulse and the p1 pulse from the junk exit point are well approximated by ϕ'. Then $t_{\Delta p}$ is approximated as (Fig. 5)

$$t_{\Delta p} \cong \sin(\phi')dz/c \qquad(18)$$

in which c is the speed of sound through water (c_w), for $-90° \leq \alpha' \leq 90°$, but c is the variable speed of sound through whale tissue (c_t), for $\alpha' > 90°$ or $\alpha' < -90°$. The change in sound speed is necessary because for clicks propagating forward ($-90° \leq \alpha' \leq 90°$) both the p0 path, and the p1 path after exiting the junk, pass primarily through the water, while backward propagating pulses pass through whale tissue. For each click, c_w is found from the value of the sound speed profile interpolated to the depth of the whale, while c_t is a function of temperature, pressure, ray elevation, and ray azimuth (different angles mean that sound passes through different tissues). For simplicity, the unknown value of c_t is assumed here to be constant and is estimated in the following optimization step.

The constants t_c, dz, and c_t, and the roll for each click are found as follows. With t_c, dz, and c_t fixed over all clicks, we find the roll for each click that minimizes the difference between measured and modeled τ in a least-squares sense over all receivers. Summing over all clicks gives the total squared error associated with the current values of t_c and dz. This total error is minimized over t_c, dz, and c_t.

The best fit values were $t_c = 6.6$ ms, $dz = 1.30$ m, and $c_t = 1540$ m/s. The value $t_c = 6.6$ ms corresponds to a whale length of 14.42 m using the formulas of Gordon (1991), and to a whale length of 13.61 m using the formula of Rhinelander and Dawson (2004). The estimated p1 exit point located 1.30 m ventral of the phonic lips for a whale over 13.5 m makes sense anatomically assuming that the exit point is on the junk (Møhl, 2001). It is also consistent with the results of Zimmer et al. (2005a), who found the p1 exit point to be 1.10 m ventral of the phonic lips for a 12 m whale. The derived value $c_t = 1540$ m/s is high compared to the value of 1370 m/s found by Flewellen and Morris (1978) for the speed of sound through spermaceti oil at 33 °C at 1 atm. It is more consistent with (although still on the high end of) values for more similar conditions given by Goold et al. (1996), who found that sound speed in spermaceti oil increased from 1390 to 1540 m/s with increasing pressure (from 0 to 90 atm) and decreasing temperature (from 38 to 22 °C). The seemingly high value for c_t found here possibly stems from the fact that our animal is alive, and that the p1 pulse passes through other whale tissue (not only spermaceti oil) to get to the receiver; however, we have made numerous assumptions and approximations that invariably introduce error, and our estimate will need to be examined in
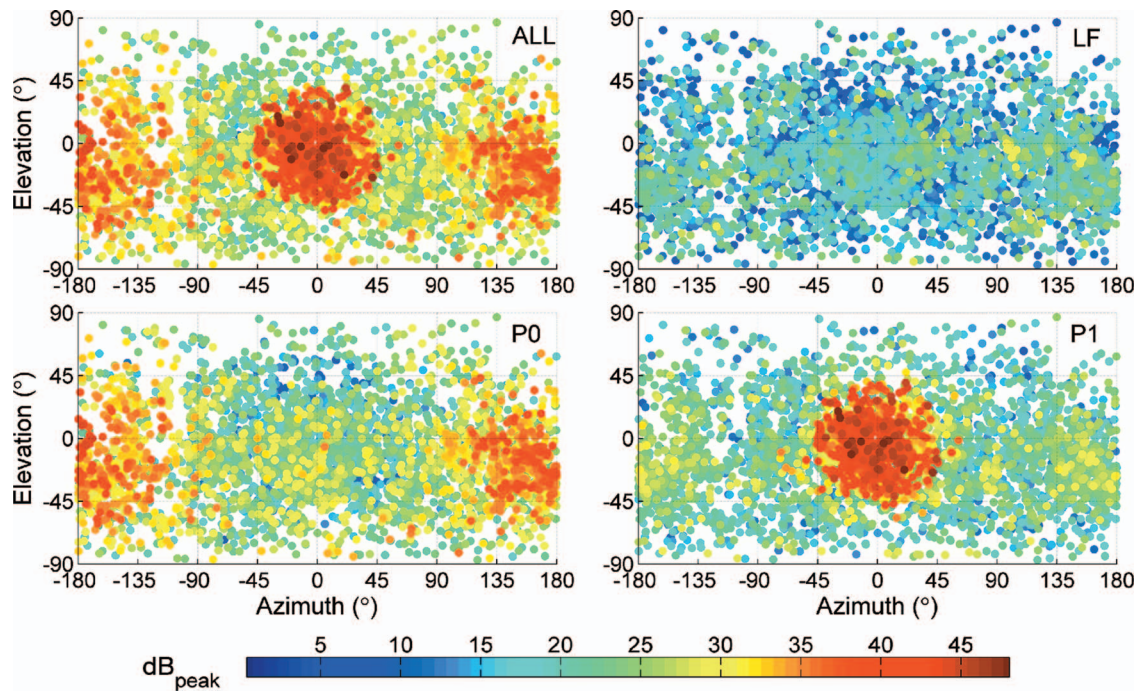
FIG. 6. Estimated beam patterns from 4807 recorded clicks. Since receiver sensitivities were not available, these are not absolute click levels but relative levels, such that 0 dB corresponds to the weakest recorded click. Recorded levels were corrected for transmission loss and were plotted as colored dots on the appropriate elevation/azimuth position. Higher levels are shown overlapping lower levels to minimize the effect of a variable source level. Results are shown for the full click, p0 pulse, p1 pulse, and LF components.

future work.

V. BEAM PATTERN AND DIRECTIVITY

A. Methods and results

The azimuth and elevation of each click to each receiver were found for the calculated position and orientation data as outlined in Sec. IV. The received levels obtained in Sec. II B were corrected for transmission loss using the values in the lookup table from Sec. III A to get click levels. Since hydrophone sensitivity (or clipping level) was unavailable, click levels could only be found as values relative to some arbitrary level, chosen such that the weakest click level corresponded to 0 dB. Hence, we report only "relative click levels," by which we mean the difference between the current click level and the minimum click level (over all clicks).

Relative click levels are plotted as a function of azimuth and elevation in Fig. 6. Since 324, 480, and 383 clicks were localized on three, four, and five receivers, respectively, a total of $324 \times 3 + 480 \times 4 + 383 \times 5 = 4807$ points are plotted. Multiple clicks with similar azimuth and elevation were measured, and the figures show higher levels overlapping lower levels, which helps to reduce the effect of variation in click source levels by approximating the maximum level in each direction. The resulting beam patterns are similar, although with somewhat broader peaks, to the patterns found by Zimmer et al. (2005b), who used a similar approach.

Although absolute source levels cannot be estimated, source level differences can be found: The maximum source level of the p1 pulse measured here was 8.8 dB higher than the maximum source level of the p0 pulse and 19.4 dB higher than the maximum source level of the LF component.

These values are consistent with estimates reported by Zimmer et al. (2005b) of 210 dB_{peak} for the p1 pulse, 200 dB_{peak} for the p0 pulse, and 190 dB_{peak} for the LF component, all re: 1 μPa at 1 m.

Directivity indices were estimated according to a discretized version of Eq. (3-10) of Au (1993)

$$DI = 10 \log \frac{4\pi}{\sum_{i=1}^{N_\alpha} \sum_{j=1}^{N_\phi} \left(\frac{p(\alpha_i', \phi_j')}{p_{max}} \right)^2 \cos \alpha' \Delta \alpha' \Delta \theta'},$$

in which N_α is the number of azimuth steps of width $\Delta \alpha'$, N_ϕ is the number of elevation steps of width $\Delta \phi'$, $p(\alpha_i', \phi_j')$ is the received pressure corrected for transmission loss for the bin corresponding to azimuth step α_i' and elevation step ϕ_j' (recall that the primes denote whale frame coordinates), and p_{max} is the maximum received pressure over all angles. Step widths of 2.5° were used for both azimuth and elevation, and the maximum pressure over each bin was used for $p(\alpha_i', \phi_j')$. Estimated directivity indices were 21.8 dB for the p1 pulse, 9.4 dB for the p0 pulse, and 5.2 dB for the LF component. In comparison, Møhl et al. (2003) reported a p1 directivity index of 27 dB and Zimmer et al. (2005b) reported a p1 directivity index of 26.7 dB and a p0 directivity index of 7.4 dB.

B. Discussion

Similar to the case discussed in Zimmer et al. (2005b), it is likely that the maximum source level of the p1 pulse is underestimated here due to clipping of the high-intensity arrivals (197 out of all 4807 signals used reached clipping amplitude), limited sampling bandwidth, and a small sample size of on-axis clicks (only 8 clicks within 5° of the main

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

E.-M. Nosal and L. N. Frazer: Sperm whale tracking   1975
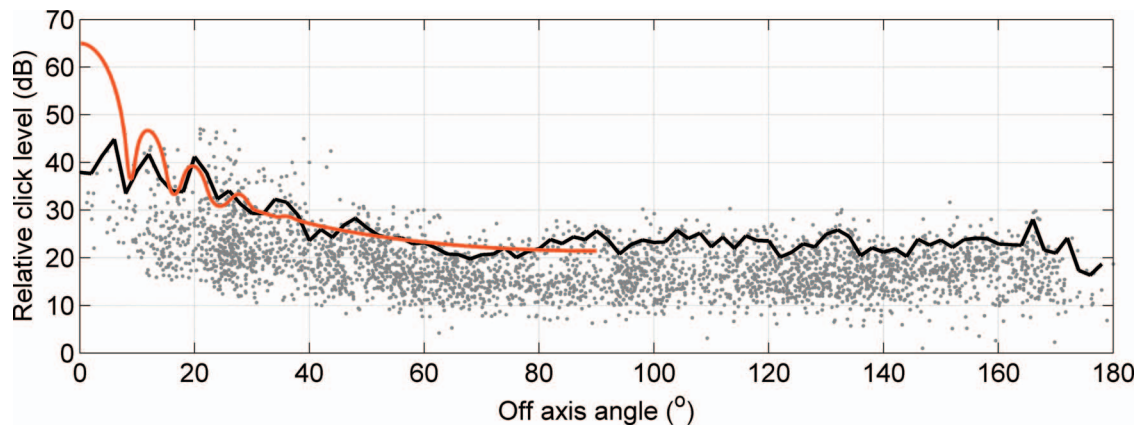
FIG. 7. Scatter plot of estimated p1 beam pattern as a function of off-axis angle. Levels are relative, as in Fig. 6. The black line represents the 90th percentile for each off-axis angle bin (bin size 2°). The red line represents the beam pattern predicted for a circular piston with parameters fitted to measured values (black line) for off-axis angles between 20° and 90°. Results and best-fit parameters are given in the text.

beam axis). This explains in part why our beam pattern for the p1 pulse has a broader peak and lower directivity than reported by Zimmer *et al.* (2005b) and Møhl *et al.* (2003).

To deal with clipping, we follow Zimmer *et al.* (2005b), in which a model broadband beam pattern is fit to the 90th percentile for off-axis angles between 20° and 90° (binned here into 2° intervals). In Fig. 7, which is similar to Fig. 9 of Zimmer *et al.* (2005b), all measured (relative) levels are plotted with gray dots as a function of off-axis angle. The 90th percentile for each (2°) off-axis bin is plotted in black and the best-fit modeled beam pattern is plotted in red. The p1 pulse was modeled as a Gabor function emitted from a circular piston (Au, 1993). The parameters that give the best fit (in a least-squares sense) are peak frequency 15 kHz, signal duration 0.60 ms, and piston radius 0.40 m. These correspond to a p1 directivity index of 25.2 dB. This is much closer to, although still considerably less than, the values of 27 and 26.7 dB reported by Møhl *et al.* (2003) and Zimmer *et al.* (2005b), respectively. Unfortunately, maximum source level could not be estimated here because hydrophone sensitivity was unknown.

Other sources of error in our beam pattern are from errors in estimated pitch, yaw, and roll. These stem from uncertainties in source location and click time, from the assumption that the whale's main axis coincides with the velocity vector, and from approximations made in application of the bent-horn model (constant sound speed through whale tissue, for example). Further, our results are limited by the assumption that all clicks have the same beam pattern and by the assumptions on receiver response (flat-frequency, omnidirectional, and the same for all receivers). The remarkable agreement of our beam patterns with those of Zimmer *et al.* (2005b) and Møhl *et al.* (2003), both of which used very different experimental setups, suggests that either our assumptions are at least approximately satisfied or that the errors caused by the assumptions tend to cancel over multiple hydrophones.

VI. CLICK SOURCE LEVELS

Click source levels were estimated from the measured beam patterns by finding best-fit levels to a "model" beam pattern. For a given direction, the model beam pattern was assigned the maximum received level, corrected for transmission loss, over all directions within 5°. This approach was preferred over binning the received levels into discretized azimuth and elevation steps, which would have resulted in a nonuniform weighting of the received levels since different elevation bins subtend different solid angles (this also explains why the clicks in Fig. 6 are more densely populated at elevations closer to 0°). For a sufficient number of clicks, this should eliminate variations due to click source level, giving a model that well approximates the true beam pattern (Zimmer *et al.* 2005b).

Relative click source levels (at a distance of 1 m on the acoustic axis) were estimated by minimizing the misfit between the model levels and the received levels corrected for transmission loss and source level. Minimization was done in a least-squares sense over all receivers that heard the click. Again, because receiver sensitivity was unknown, only relative click source levels could be found; the resulting relative click source levels are shown in Fig. 8. A total of 14 complete click series and 2 incomplete click series (at the beginning and the end of the data set) were recorded, where a series is defined as ending in a creak or at least 5 s of silence. Click levels vary by about 20 dB [in agreement with the dynamic range reported by Madsen *et al.* (2002)] and tend to steadily decrease toward the end of each click series. There were no apparent correlations between the interclick intervals and source levels, or between whale depth (or orientation) and source level. However, as shown in Fig. 9, there is a significant relationship between click level and the order of
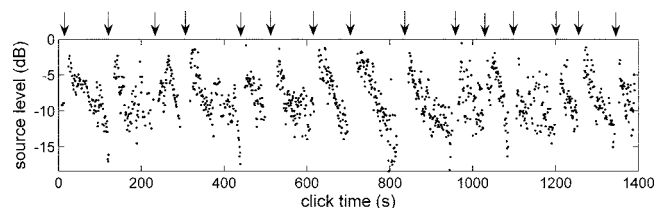
FIG. 8. Source levels relative to the strongest recorded source level as a function of time. Click levels decline by 10–15 dB from the start to the end of most click series. The beginning of each series is indicated by an arrow at the top.
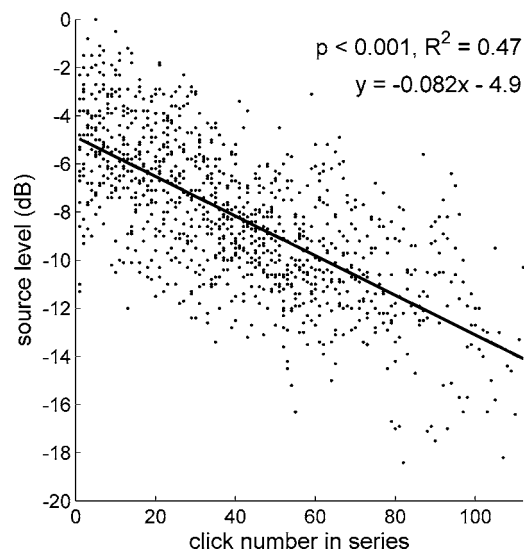
FIG. 9. Relative source level as a function of click number within its series. Data are pooled from all 14 complete click series. The significant correlation and the negative slope of the regression line suggest that click level decreases with click number within a series.

the click within its series. This suggests that the variation in click level may be a consequence of the click production mechanism, whereby a click series begins at some constant level and decreases with each subsequent click. However, other explanations are possible. For example, since we do not have target range information it is not possible to determine if level is controlled by some automatic gain control mechanism, as might be employed by dolphins (Au and Benoit-Bird, 2003; Au and Herzing, 2003) but possibly not by beaked whales (Madsen *et al.*, 2005).

Recalculating the directivity indices after correcting with these click source levels gave indices of 22.9, 9, and 5 dB for the p1, p0, and LF components, respectively. Since corrected beam patterns were very similar to those in Fig. 6 they are not presented.

VII. CONCLUDING REMARKS

Although our method to recover roll is specific to sperm whales, our estimation of pitch and yaw is applicable to any clicking marine mammal. Since beam patterns often exhibit rotational symmetry, at least to a first approximation—as for sperm whales (Zimmer *et al.* 2005b; this paper) and for bottlenose dolphins (Au, 1993)—it may be useful to estimate beam patterns as a function of off-axis angle only. In that case, roll is not needed, so the methods developed here can be used to obtain directivity indices and estimates of click level for any clicking marine mammal recorded on bottom-mounted hydrophones in the wild, provided that the clicks are heard on enough hydrophones to produce 3D locations.

Abraham, D. A. (**2000**). "Passive acoustic detection of marine mammals," SM-351, SACLANT Undersea Research Centre, La Spezia, Italy.

Adam, O., Motsch, J.-F., Desharnais, F., DiMarzio, N., Gillespie, D., and Gisiner, R. C. (**2006**). "Overview of the 205 workshop on detection and localization of marine mammals using passive acoustics," Appl. Acoust. **67**, 1061–1070.

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

Au, W. W. L., and Benoit-Bird, K. (**2003**). "Automatic gain control in the echolocation system of dolphins," Nature (London) **423**, 861–863.

Au, W. W. L., and Herzing, D. L. (**2003**). "Echolocation signals of wild Atlantic spotted dophin (*Stenella frontalis*)," J. Acoust. Soc. Am. **113**, 598–604.

Bracewell, R. N., (**2000**). "The Hilbert transform," in *The Fourier Transform and Its Applications*, 3rd ed. (McGraw-Hill, Singapore), pp. 359–364.

Flewellen, C. G., and Morris, R. J. (**1978**). "Sound velocity measurements on samples from the spermaceti organ of the sperm whale *Physeter catodon*," Deep-Sea Res. **25**, 269–277.

Giraudet, P., and Glotin, H. (**2006**). "Real-time 3D tracking of whales by precise and echo-robust TDOAs of clicks extracted from 5 bottom-mounted hydrophones records of the AUTEC," Appl. Acoust. **67**, 1106–1117.

GDEM (Generalized Digital Environment Model) website (last accessed on 1 May, 2007), ⟨https://128.160.23.42/gdemv/gdemv. html⟩.

Goldstein, H. (**1980**). "The Euler angles" and "Euler angles in alternate conventions," in *Classical Mechanics*, 2nd ed. (Addison-Wesley, Reading, MA), Sec. 4–4, pp. 143–148 and Appendix B, pp 606–610.

Goold, J. C., Bennell, J. D., and Jones, S. E. (**1996**). "Sound velocity measurements in spermaceti oil under the combined influences of temperature and pressure," Deep-Sea Res., Part I **43**, 961–969.

Goold, J. C., and Jones, S. E. (**1995**). "Time and frequency domain characteristics of sperm whale clicks," J. Acoust. Soc. Am. **98**, 1279–1291.

Gordon, J. C. (**1987**). "The behaviour and ecology of sperm whales off Sri Lanka," Ph.D. thesis, University of Cambridge, Cambridge, UK.

Gordon, J. C. D. (**1991**). "Evaluation of a method for determining the length of sperm whales (*Physeter catodon*) from their vocalizations," J. Zool. **224**, 301–314.

Jaquet, N., Dawson, S., and Douglas, L. (**2001**). "Vocal behavior of male sperm whales: Why do they click?," J. Acoust. Soc. Am. **109**, 2254–2259.

Johnson, M., and Tyack, P. L. (**2003**). "A digital acoustic recording tag for measuring the response of wild marine mammals to sound," IEEE J. Ocean. Eng. **28**, 3–12.

Laplanche, C., Adam, O., Lopatka, M., and Motsch, J.-F. (**2006**). "Measuring the off-axis angle and the rotational movements of phonating sperm whales using a single hydrophone," J. Acoust. Soc. Am. **119**, 4074–4082.

Madsen, P. T. (**2002**). in "Sperm whale sound production," Ph.D. thesis, University of Aarhus, Denmark.

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

Madsen, P. T., Johnson, M., Aguilar deSoto, N., Zimmer, W. M. X., and Tyack, P. (**2005**). "Biosonar performance of foraging beaked whales (*Mesoplodon densirostris*)," J. Exp. Biol. **208**, 181–194.

Madsen, P. T., Payne, R., Kristiansen, N. U., Wahlberg, M., Kerr, I., and Møhl, B. (**2002a**). "Sperm whale sound production studied with ultrasound time/depth-recording tags," J. Exp. Biol. **205**, 1899–1906.

Madsen, P. T., Wahlberg, M., and Møhl, B. (**2002b**). "Male sperm whale (*Physeter macrocephalus*) acoustics in a high-latitude habitat: Implications for echolocation and communication," Behav. Ecol. Sociobiol. **53**, 31–41.

Miller, P. J. O., Johnson, M. P., and Tyack, P. L. (**2004**). "Sperm whale behaviour indicates the use of echolocation click buzzes 'creaks' in prey capture," Proc. R. Soc. London, Ser. B **271**, 2239–2247.

Møhl, B. (**2001**). "Sound transmission in the nose of the sperm whale *Physeter catodon*. A postmortem study," J. Comp. Physiol. [A] **187**, 335–340.

Møhl, B., Wahlberg, M., Madsen, P. T., Heerfordt, A., and Lund, A. (**2000**). "Sperm whale clicks: Directionality and source level revisited," J. Acoust. Soc. Am. **114**, 1143–1154.

Møhl, B., Wahlberg, M., Madsen, P. T., Heerfordt, A., and Lund, A. (**2003**). "The monopulsed nature of sperm whale clicks," J. Acoust. Soc. Am. **114**, 1143–1154.

Morrissey, R. P., Ward, J., DiMarzio, N., Jarvis, S., and Moretti, D. J. (**2006**). "Passive acoustic detection and localization of sperm whales (*Physeter macrocephalus*) in the tongue of the ocean," Appl. Acoust. **67**, 1091–1105.

Mullins, J., Whitehead, H., and Weilgart, L. (**1988**). "Behaviour and vocalizations of two sperm whales, Physeter macrocephalus, off Nova Scotia," Can. J. Fish. Aquat. Sci. **45**, 1736–1743.

Norris, K. S., and Harvey, G. W. (1972). "A theory for the function of the spermaceti organ of the sperm whale (*Physeter catodon L.*)," in *Animal Orientation and Navigation*, edited by S. R. Galler, K. Schmidt-Koenig, G. J. Jacobs, and R. E. Belleville, NASA Special Publication No. **262**, 397–417.

Nosal, E.-M. and Frazer, L. N. (**2006**). "Track of a sperm whale from delays between direct and surface-reflected clicks," Appl. Acoust. **67**, 1187–1201.

Page, S. E. (**1954**). "Continuous inspection schemes," Biometrika **41**, 100–115.

Porter, M. (**2005**). "BELLHOP" (a Gaussian beam/finite element beam code). Available in the *Acoustics Toolbox* at http://www.hlsresearch.com/oalib/Modes/AcousticsToolbox/, Last accessed 9/18/2007.

Rhinelander, M. Q., and Dawson, S. M. (**2004**). "Measuring sperm whales from their clicks: Stability of interpulse intervals and validation that they indicate whale length," J. Acoust. Soc. Am. **115**, 1826–1831.

Spiesberger, J. L., and Wahlberg, M., (**2002**). "Probability density functions for hyperbolic and isodiachronic locations," J. Acoust. Soc. Am. **112**, 3046–3052.

Wahlberg, M., Møhl, B., and Madsen, P. T. (**2001**). "Estimating source position accuracy of a large-aperture hydrophone array for bioacoustics," J. Acoust. Soc. Am. **109**, 397–406.

Wald, A. (**1947**). *Sequential Analysis* (Wiley, New York).

Watkins, W. A., Daher, M. A., DiMarzio, N. A., Samuels, A., Wartzok, D.,

Fristrup, K. M., Howey, P. W., and Maiefski, R. R. (**2002**). "Sperm whale dives tracked by radio tag telemetry," Marine Mammal Sci. **18**, 55–68.

Watwood, S. L., Miller, P. J. O., Johnson, M., Madsen, P. T., and Tyack, P. L. (**2006**). "Deep-diving foraging behaviour of sperm whales (*Physeter macrocephalus*)," J. Anim. Ecol. **75**, 814–825.

White, P. R., Leighton, T. G., Finfer, D. C., Prowles, C., and Baumann, O. (**2006**). "Localisation of sperm whales using bottom-mounted sensors," Appl. Acoust. **67**, 1074–1090.

Whitehead, H. (**2003**). *Sperm Whales: Social Evolution in the Ocean* (Chicago University Press, Chicago).

Whitehead, H., and Weilgart, L. (**1990**). "Click rates from sperm whales," J. Acoust. Soc. Am. **87**, 1798–1806.

Whitehead, H., and Weilgart, L. (**1991**). "Patterns of visually observable behavior and vocalizations in groups of female sperm whales," Behaviour **118**, 275–296.

Zimmer, W. M. X., Johnson, M. P., D'Amico, A., and Tyack, P. L. (**2003**). "Combining data from a multisensor tag and passive sonar to determine the diving behavior of a sperm whale (*Physeter macreocephalus*)," IEEE J. Ocean. Eng. **28**, 13–28.

Zimmer, W. M. X., Madsen, P. T., Teloni, V., Johnson, M. P., and Tyack, P. L. (**2005a**). "Off-axis effects on the multipulse structure of sperm whale usual clicks with implications for sound production," J. Acoust. Soc. Am. **118**, 3337–3345.

Zimmer, W. M. X., Tyack, P. L., Johnson, M. P., and Madsen, P. T. (**2005b**). "Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis," J. Acoust. Soc. Am.. **117**, 1473–1485.

Robust matched-field processing using a coherent broadband white noise constraint processor

Claire Debever[a)] and W. A. Kuperman[b)]

Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 90093

Adaptive matched-field processing (MFP) is not only very sensitive to mismatch, but also requires the received sound levels to exceed a threshold signal-to-noise ratio. Furthermore, acoustic sources and interferers have to move slowly enough across resolution cells so that a full rank cross-spectral density matrix can be constructed. Coherent-broadband MFP takes advantage of the temporal complexity of the signal, and therefore offers an additional gain over narrow-band processing by augmenting the dimension of the data space. However, the sensitivity to mismatch is also increased in the process, since a single constraint is usually not enough to achieve robustness and the snapshot requirement becomes even more problematic. The white noise constraint method, typically used for narrow-band processing, is applied to a previously derived broadband processor to enhance its robustness to environmental mismatch and snapshot deficiency. The broadband white noise constraint theory is presented and validated through simulation and experimental data. The dynamic range bias obtained from the snapshot-deficient processing is shown to be consistent with that previously presented in the literature for a single frequency.
© *2007 Acoustical Society of America.* [DOI: 10.1121/1.2769830]

I. INTRODUCTION

Adaptive matched-field processing (MFP) is extremely sensitive to environmental mismatch.[1,2] For example, the minimum variance distortionless response (MVDR) method of using a borehole constrained optimization requires accurate environmental model parameters and array element localization. The procedure also involves a matrix inversion of a well-conditioned sample cross-spectral density matrix (CSDM). Single constraints, leading to high resolution processors, become problematic for a combination of large arrays with some mismatch.[3] Robust processors attempt to overcome the mismatch problem while keeping the benefits of the adaptive algorithms.[4–7] The white noise constraint method (WNCM)[7] has been localizing sources in many challenging environments without the requirement of seeking more accurate environmental information. Further, it has been effective in dealing with so-called "snapshot-deficient scenarios" in which targets and interferors move across resolution cells before enough data vectors, or snapshots, can be recorded at the array and combined to construct a full rank CSDM.[8,9]

Adaptivity also requires a threshold signal-to-noise ratio (SNR).[10] Therefore, coherent broadband MFP, which involves an extra-processing gain from using multiple frequency data analogous to the coherent gain associated with more array elements, potentially offers additional opportunities over single frequency processing. However, we can expect adaptive coherent broadband MFP to have even greater mismatch sensitivity since there is a larger data space (frequency × number of array elements) subject to the same

single constraint. In addition, the snapshot requirement[8] for constructing the larger CSDM, which is typically already strained for practical single frequency scenarios, becomes problematic. In this paper, we apply the white noise constraint method to a coherent broadband processor to develop a robust[7] coherent broadband MFP processor suitable for snapshot-deficient scenarios.

Broadband matched-field methods can be implemented in either the frequency or time domain. In the time domain, for example, Clay[11] proposed to match measured and modeled impulse responses of transmission between the source and a receiver. This technique, which requires knowledge of the source function, was used by Li *et al.* on laboratory data.[12,13] Brienzo and Hodgkiss[14] also successfully applied it on experimental data coming from an explosive source. The unknown source wave form was estimated from the direct path arrival, and the additional processing gain it procures demonstrated. Frazer and Pecholcs[15] generalized the method and introduced new algorithms requiring no knowledge of the source function, instead estimating it as well as the source location. More recently, Hursky *et al.*[16] tracked a high frequency source in range and depth by matching correlations of impulse response function between receivers, canceling out by design the unknown source wave form.

In the frequency domain, a common approach proposed by Baggeroer *et al.* is to incoherently average single-frequency ambiguity surfaces to suppress ambiguous sidelobes.[10] Coherent broadband processors, in contrast, take full advantage of the complexity of the signal by keeping the complex cross-frequency inner products. These processors have to deal with the typically unknown complex spectral properties of the source and can be divided into two categories:

---

a)Electronic mail:cdebever@ucsd.edu
b)Electronic mail:wak@mpl.ucsd.edu

(1) Matching a first moment quantity such as pressure that consequently involves sample data covariance matrices.

(2) Matching a second moment quantity such as data correlations between receiver element pairs and therefore involves estimating fourth moment quantities from data (with an increased data requirement from [frequency × number of array elements] to [frequency × (number of array elements)^2] for accurate estimation).

Actually, the latter category was first treated in the literature by Westwood, who developed an algorithm which sums correlations of modeled and experimental cross spectra coherently over frequency.[17] No knowledge of the source spectrum is assumed, but it can be easily incorporated in the algorithm when known. The method was applied successfully to a shallow water experimental data set characterized by strong multipaths.[18] This method was further generalized by Czenszak and Krolik,[19] who introduced an algorithmic procedure involving the fourth-order statistics of the data so that the cross-spectral density matrix of the received data is matched to a set of weight matrices using the magnitude squared Frobenius matrix inner product. Their purpose was to combine the higher moment analysis with its large data and computational requirements with the earlier developed environmental perturbation constraint method.[5] The latter's additional constraints resulted in a more stable, though extremely lengthy processing procedure.

The first category builds on standard algorithms already well known from frequency domain processing. Michalopoulou and Porter[20,21] proposed a processor in which the data is stacked over frequencies in a "supervector" and matched to a modeled one. While the transfer functions are treated coherently, one should note that this processor is really only semi-coherent, since the unknown phases across frequency are removed in the construction of the supervector. The processing allows for including the source amplitude, if known. Orris et al. further considered those unknown phase relationships as free parameters that may be determined using a global functional minimization algorithm, again with increased computational complexity.[22]

In parallel, an extensive amount of work has been done to address the conflicting issues of resolution and robustness to environmental mismatch. High resolution was achieved by developing optimal processors resulting from the solution of a constrained optimization problem. These adaptive algorithms, such as the the MVDR processor,[23] use the data to construct optimum weights. While very efficient at suppressing noise and interfering signals, the MVDR is more vulnerable to any kind of mismatch between the modeled and real pressure fields than a conventional processor.[7] Furthermore, it requires a sufficient number of snapshots to construct an invertible CSDM. This is often problematic for a moving target because of the opposing requirements of high resolution producing small resolution cells versus the necessity of building up a CSDM from sufficient snapshots in a single cell.

Again there are two categories of dealing with environmental robustness:

(1) Enlarge the search space or provide some sort of environmental blurring,[4–6,24]

(2) Constrain the algorithmic cause of the instability.[3,7]

The WNCM,[7] which constrains the array gain against uncorrelated noise, performs particularly well in dealing with both robustness and snapshot-deficient data. For the latter aspect, the dynamic range bias of snapshot-deficient processing has been addressed in the literature[8,9] and must also be considered in the broadband case.

This paper combines the coherent broadband method introduced by Michalopoulou and Porter (herein referred to as MP) with white noise constrained MVDR processing. In Sec. II, we first provide an overview of incoherent and coherent MP processors. We then apply the MVDR/WNCM in Sec. III. We also illustrate the resulting performance using data from the Hudson Canyon experiment.[25] In Sec. IV, we show that the dynamic range bias from the snapshot-deficient processing is consistent with previous single frequency literature[8,9] and that the processor has a SNR threshold. We evaluate the robustness of the processor to slight environmental mismatch as the source moves in range in Sec. V. Finally, Sec. VI provides a summary and discussion.

II. FORMULATION OF THE COHERENT MP MODEL

Amongst the diverse coherent broadband algorithms introduced in the literature, the coherent MP model has been chosen in this study because it is readily applicable to adaptive MFP, and the results obtained using the MVDR processor on real data are promising.[20]

Let $x_f^{(i)}$ denote the complex field received on the hydrophone i, N the number of hydrophones in the array, and L the number of individual frequencies considered. The multiple frequency components are incorporated by stacking narrowband signal vectors into one "supervector" of length $N \times L$, L being the number of tones processed and the superscript T the transpose operation,

$$\hat{x} = [x_{f_1}^1, x_{f_1}^2, \ldots, x_{f_1}^N, \ldots, x_{f_L}^1, \ldots, x_{f_L}^N]^T. \tag{1}$$

From this supervector, a "super" CSDM can be constructed by forming the outer product of the vector by its complex conjugate transpose, $\hat{R} = E\{\hat{x}\hat{x}^\dagger\}$, where $E\{\ \}$ is the expectation value. The MFP processor output can then be formulated as

$$P_{\text{MP}}(r,z) = \hat{d}^\dagger(r,z)\hat{R}\hat{d}(r,z),$$

$$P_{\text{MP}}(r,z) = E\left\{ \left| \sum_B d_f^\dagger(r,z) x_f \right|^2 \right\}, \tag{2}$$

where $d_f(r,z)$ is the replica vector at frequency f from a source at depth z and distance r from the array, B is the bandwidth of interest, and the dagger is the complex conjugate transpose operation.

Note that the position of the absolute value delimiters outside of the sum over frequencies allows the transfer functions to be treated coherently and generates cross-frequency terms. In contrast, an incoherent algorithm sums absolute values of inner products

C. Debever and W. A. Kuperman: Coherent broadband white noise constraint

$$P_{\text{Incoh}}(r,z) = E\left\{\sum_B |\boldsymbol{d}_f^\dagger(r,z)\boldsymbol{x}_f|^2\right\}. \tag{3}$$

The complex signal received on the element i can be written as

$$x_f^{(i)} = s(f)h_f^{(i)}, \tag{4}$$

where $s(f)$ is the complex source signal at the frequency f, and $h_f^{(i)}$ the transfer function between the source and the receiver i, assuming the ideal case of no contamination by noise. Rewriting the output of the algorithm reveals that the unknown source phase terms remain to be evaluated, leaving the user with replicas mismatched with respect to the signal,

$$\begin{aligned} P_{\text{MP}}(r,z) &= \sum_{l=1}^{L}\sum_{k=1}^{L} \boldsymbol{d}_l^\dagger(r,z)\boldsymbol{x}_l\boldsymbol{x}_k^\dagger \boldsymbol{d}_k(r,z) \\ &= \sum_{l=1}^{L}\sum_{k=1}^{L} s(l)s(k)^*\boldsymbol{d}_l^\dagger(r,z)\boldsymbol{h}_l\boldsymbol{h}_k^\dagger \boldsymbol{d}_k(r,z), \end{aligned} \tag{5}$$

where the asterisk represents the complex conjugate operation.

The MP method addresses this issue by scaling the phase of each single-frequency subvector $\boldsymbol{x}_f$ by the phase on the first phone, and normalizing the subvector to unit length.[20,21] This compensation procedure eliminates the unknown source terms in high signal-to-noise scenarios since, if we denote the new subvector by $\bar{\boldsymbol{x}}_f$,

$$\boldsymbol{x}_f = \begin{bmatrix} |s(f)h_f^{(1)}|e^{j(\arg(s(f))+\arg(h_f^{(1)}))} \\ \vdots \\ |s(f)h_f^{(N)}|e^{j(\arg(s(f))+\arg(h_f^{(N)}))} \end{bmatrix},$$

we have

$$\bar{\boldsymbol{x}}_f = \frac{1}{\alpha}\begin{bmatrix} |h_f^{(1)}| \\ \vdots \\ |h_f^{(N)}|e^{j\left(\arg(h_f^{(N)})-\arg(h_f^{(1)})\right)} \end{bmatrix}, \tag{6}$$

where $\alpha = \sqrt{|h_f^{(1)}|^2 + \cdots + |h_f^{(N)}|^2}$ is a coefficient of normalization.

When the amount of noise embedded in the signal is not negligible anymore, the above-described amplitude and phase normalization scheme does not cancel the unknown source terms in the noise component. The MP algorithm therefore carries an inherent mismatch when localizing low signal-to-noise ratio sources. See Sec. IV for further discussion of the MP algorithm's performance versus signal-to-noise ratio.

III. MINIMUM VARIANCE AND WHITE NOISE CONSTRAINT PROCESSING

In this section, the minimum variance and white noise constraint methods are presented in the context of broadband MFP. Their localization performance, both in the incoherent and coherent case, is evaluated on experimental data. The Hudson Canyon experiment,[25] used by Michalopoulou and Porter to test their algorithms as well,[20] was chosen for comparison purposes. It took place in shallow water (73 m) out
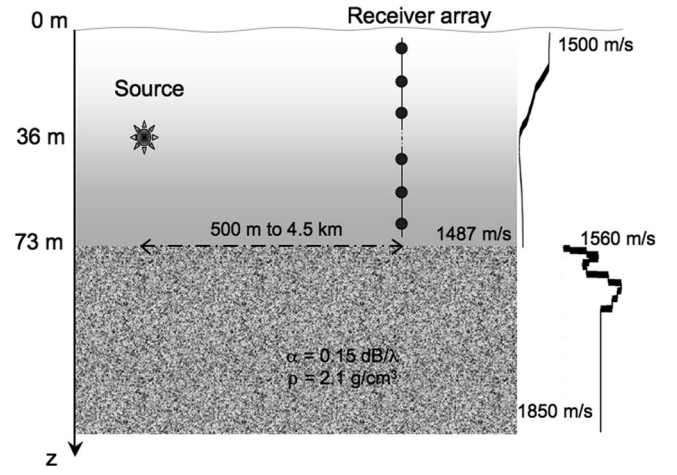
FIG. 1. The Hudson Canyon environment features—a 24 element vertical array and an acoustic source towed at 36 m deep from 500 m to 4.5 km away from the array. In a first track, the source moved away from the receivers and sent a set of four tones at 50, 175, 375, and 425 Hz. It traveled back toward the receiver array in a second track, emitting tones at 75, 275, 525, and 600 Hz.

of the New Jersey coast. An acoustic source was towed at 36 m deep over an essentially flat bottom, and the acoustic field was sampled by a 24 element vertical line array.

Data from two source tracks (ten different ranges per track) are provided. In the first one, the source sent multitones at 50, 175, 375, and 425 Hz and moved up to 4.5 km away from the receivers. The source traveled back toward the receiver array in the second track, emitting tones at 75, 275, 525, and 600 Hz. Ten observations are available for each source range, and the average SNR at each element was approximately 10 dB. The Hudson Canyon environment is illustrated in Fig. 1. A normal mode propagation model[26] is used to create the replica vectors, and 4, 15, 32, and 36 modes were kept, respectively, at the frequencies of the first leg, and 6, 24, 30, and 46 modes at the ones of the second leg.

A. Minimum variance method

The general formulation of the MVDR matched-field processor in the frequency domain starts with

$$P(r,z) = \boldsymbol{w}^\dagger(r,z)\boldsymbol{R}\boldsymbol{w}(r,z), \tag{7}$$

where $\boldsymbol{R} = E\{\boldsymbol{x}\boldsymbol{x}^\dagger\}$ is the CSDM at the frequency of interest, and the weight vector $\boldsymbol{w}$ is determined by solving

$$\min_{\boldsymbol{w}} \boldsymbol{w}^\dagger\boldsymbol{R}\boldsymbol{w} \text{ subject to } \boldsymbol{w}^\dagger\boldsymbol{d} = 1, \tag{8}$$

$\boldsymbol{d}$ being the look direction steering vector.

The well-known solution of this optimization problem is

$$\boldsymbol{w}_{\text{MVDR}} = \frac{\boldsymbol{R}^{-1}\boldsymbol{d}}{\boldsymbol{d}^\dagger\boldsymbol{R}^{-1}\boldsymbol{d}}, \tag{9}$$

which yields

$$P_{\text{MVDR}} = \frac{1}{\boldsymbol{d}^\dagger\boldsymbol{R}^{-1}\boldsymbol{d}}. \tag{10}$$

This algorithm enhances the resolution considerably in comparison to a conventional processor; however, this exceptional resolution capability comes with an increased sensitivity to slight mismatch between the modeled and actual environment.

Michalopoulou and Porter tested the MVDR algorithm on the Hudson Canyon experimental data set.[20] Since only ten snapshots were available per source position, the cross-spectral density matrix was diagonally loaded "by adding a small multiple of the identity matrix to it."[20] Their study shows that incoherently averaging MVDR ambiguity surfaces yields marginal results for that data set. As expected, the narrow-band MVDR is quite sensitive to mismatch, and averaging single-frequency outputs reinforces the maximum constructively in only 10% of the cases (30% when each output was normalized to a maximum of one prior averaging).

In contrast, the broadband coherent MP version of the MVDR shows a great improvement, now successfully localizing the source in 90% of the cases, though with a peak-to-sidelobe ratio of only 2 dB or less.

These results show that frequency coherence does enhance marginally the estimation performance, and motivate us to apply an additional constraint on the white noise gain (WNCM processor) in an attempt to increase the robustness to mismatch.

B. White noise constraint method

Cox et al.[7] showed that in the presence of uncorrelated noise, the sensitivity of the algorithm to signal mismatch is equal to the reciprocal of the white noise gain. This observation led them to introduce an inequality constraint on the gain against spatially white noise. This modified MVDR, the WNCM algorithm, consequently loses some of its high-resolution characteristics along with its requirement for very precise knowledge of the environment.

In the frequency domain, the weight vector is the solution of the optimization problem:

$$\min_{w} \ w^{\dagger}Rw \ \text{subject to} \ w^{\dagger}d = 1 \ \text{and} \ w^{\dagger}w \leqslant \delta^{-2}, \quad (11)$$

where δ^2 is a number less than the maximum possible white noise gain $\max[1/(w^{\dagger}w)] = 1$. It yields

$$w_{\text{WNCM}} = \frac{(R + \epsilon I)^{-1}d}{d^{\dagger}(R + \epsilon I)^{-1}d}. \quad (12)$$

The value of ϵ is such that the inequality of Eq. (11) is satisfied.

When the environment is well studied, single frequency WNCM usually succeeds in localizing the source. An incoherent average across frequencies is then expected to further reduce the sidelobes level, provided that the type of mismatch present does not provoke a frequency-dependent shift of the maximum.

In an attempt to decrease the level of the sidelobes, a coherent-broadband white noise constraint processor has been implemented. Since the MP algorithm is formulated in the usual quadratic form, $P_{\text{MP}}(r,z) = \hat{w}^{\dagger}(r,z)\hat{R}\hat{w}(r,z)$, and as
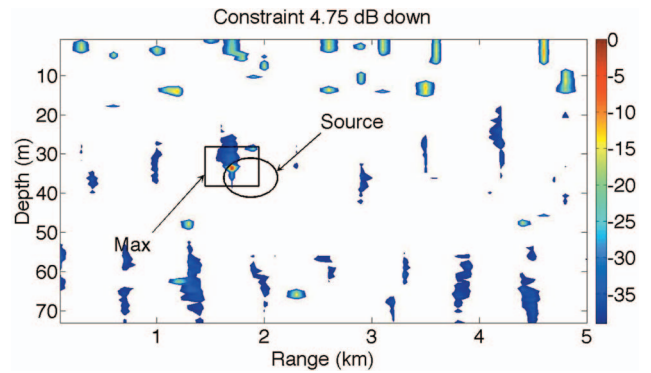
FIG. 2. White noise constraint MFP obtained by an incoherent decibel average of four frequencies ambiguity surfaces (50, 175, 375, and 425 Hz). The source is localized and the first sidelobes appear 10 dB down from the main peak. The white color corresponds to a level beyond the dynamic range.

in the narrow-band case the white noise cross-spectral density matrix is the identity matrix, the extension to the white noise constraint algorithm is straightforward:

$$w_{\text{WNCM}}^{\text{MP}} = \frac{(\hat{R} + \epsilon I)^{-1}\hat{d}}{\hat{d}^{\dagger}(\hat{R} + \epsilon I)^{-1}\hat{d}}. \quad (13)$$

This time the amount of loading is determined by the super-vector.

The same portion of the Hudson canyon data set used by Michalopoulou and Porter to test the MVDR version of their coherent algorithm is processed for comparison purposes. The value of the constraint on white noise being somewhat arbitrary, a set of constraints ranging from 0.5 to 6 dB down the maximum by steps of 0.25 dB will be applied. As discussed in Sec. IV, the constraint allowing the maximum dynamic range bias to be visible will be kept. Note that the value of the constraint does not vary across frequency. The source is considered correctly localized when the the maximum is within 5 m depth and 200 m range from the true position. Figure 2 shows the output of the incoherent average of WNCM ambiguity surfaces obtained with the constraint showing a maximum bias. The source location, at $z_s = 36$ m and $r_s = 1.87$ km, is marked by an ellipse, while the maxi-
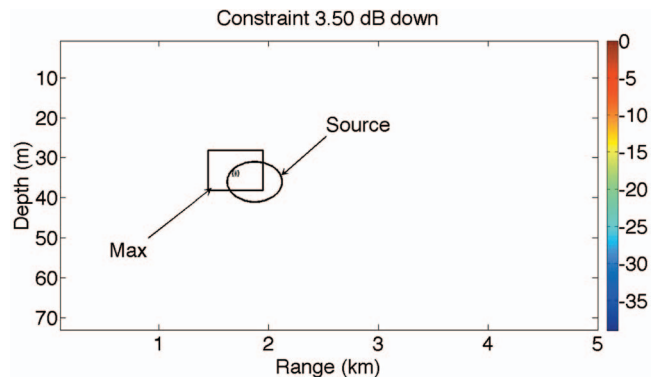
FIG. 3. Output from the MP coherent broadband white noise constraint MFP using four frequencies (50, 175, 375, and 425 Hz). The main peak is at the true source position, and stands 146 dB above the noise. The white color corresponds to a level beyond the dynamic range. The dynamic range is set to 40 dB for comparison purposes with Fig. 2.
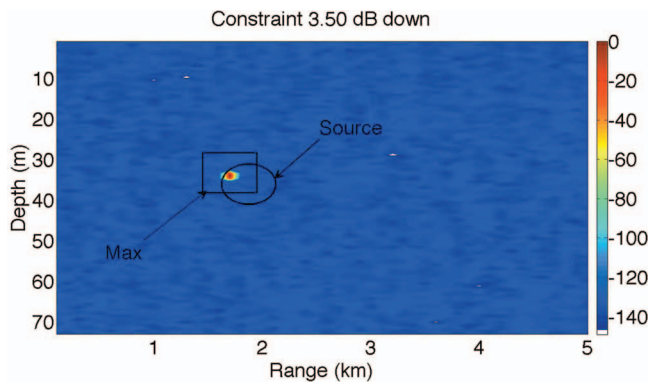
FIG. 4. Output from the MP coherent broadband white noise constraint MFP using four frequencies (50, 175, 375, and 425 Hz). The main peak is at the true source position, and stands 146 dB above the noise. The white color corresponds to a level beyond the dynamic range. The dynamic range is set to 150 dB to see the sidelobe level.

mum of the ambiguity surface lies in the square. In contrast with the incoherent average of MVDR single-frequency outputs, the source is now localized successfully, with a peak-to-sidelobe level of approximately 10 dB.

Figures 3 and 4 displays the MP coherent-broadband, white noise constraint, ambiguity surface (with the constraint showing a maximum bias) with 40 and 150 dB dynamic range, respectively. As apparent in Figs. 3 and 4, the algorithm is not only robust to mismatch between the experimental and modeled fields since the source is successfully localized, but also discriminating, as suggested by the absence of any sidelobes up to 40 dB under the peak value. In fact, a large dynamic range of 146 dB was necessary to see the first sidelobes. Since the actual SNR at each element was around 10 dB and there were only 24 elements and 4 frequencies processed, it suggests the presence of a bias in dynamic

range, discussed for a single frequency by Baggeroer and Cox[8] as well as Song et $al.$[9] and generalized for multiple frequencies in Sec. IV

IV. DYNAMIC RANGE BIAS IN WHITE NOISE PROCESSING

In a snapshot-deficient scenario, as in the Hudson Canyon data set, the CSDM eigenvalue spectrum is composed mainly of three groups. The highest eigenvalues correspond to the source and strong interferers, followed by a lower plateau of eigenvalues associated with noise, and finally non-physical eigenvalues, usually much lower in magnitude, due to the fact that the matrix is ill-conditioned. In this case, diagonally loading the CSDM decreases the gap between the noise eigenvalues and the lowest plateau, and eventually renders the matrix invertible. But by changing the magnitude of these small eigenvalues, we also change the amount of bias at the output of the loaded MVDR, such that the apparent power of the signal and noise will be lower than in reality.[8]

Song et $al.$ formulated the amount of bias to be expected versus loading γ.[9] Note that the loading referred to here as γ is a static loading independent of the look direction, as would be used to invert the CSDM for the MVDR algorithm, as opposed to the dynamic loading ϵ involved in the WNCM processor. They found that when γ was much smaller than the eigenvalues of the CSDM associated with noise, λ_n, the signal and noise power outputs shifted by $2\gamma/\lambda_n$ in decibels. However, when γ is equal or bigger to that smallest eigenvalue, the power output changes as only γ/λ_n in decibels. In both cases, the shift in magnitude of the power output is responsible for the bias in dynamic range observed on the MP coherent-broadband, white noise constraint, ambiguity surfaces, as explained in the following.

Figure 5 gives a schematic representation of the varia-
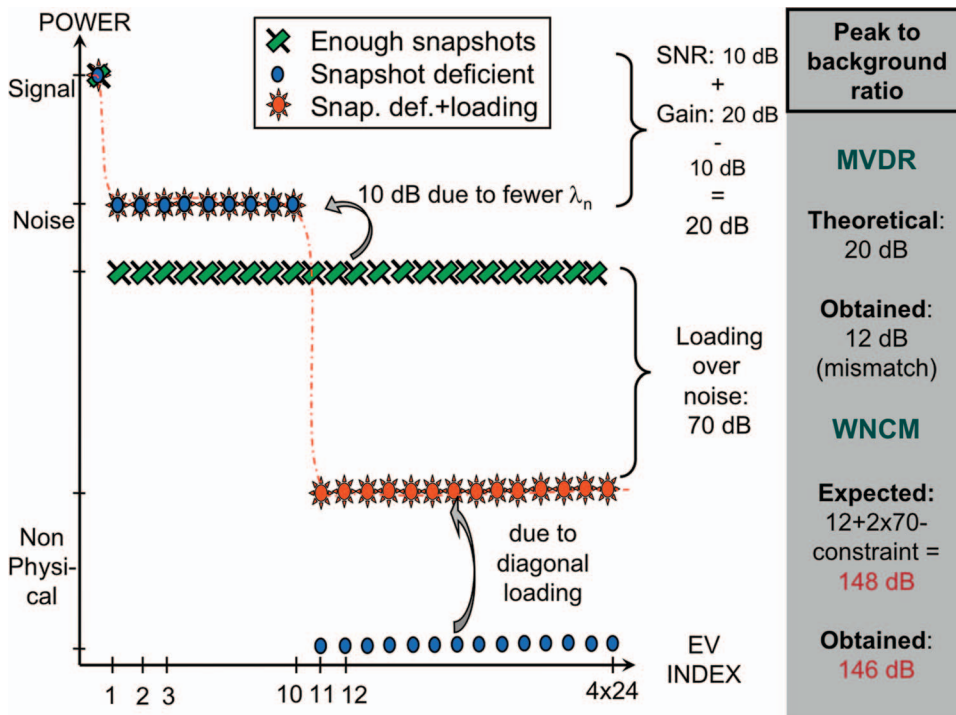
FIG. 5. Variation of the coherent CSDM signal and noise eigenvalues in the ideal case of (1) sufficient snapshots, (2) ten snapshots only, and (3) ten snapshots and a loading of γ/λ_n =−70 dB. The corresponding expected and obtained peak-to-background ratios is given for the broadband coherent MVDR and WNCM algorithms.

tion of the eigenvalue spectrum in our example as the CSDM becomes snapshot deficient. In the Hudson Canyon experiment, the SNR recorded at each element was approximately 10 dB. When introducing an additional 20 dB of array gain $[10 \log(L \times N)]$, the gap between the signal and noise eigenvalues is expected to be around 30 dB. Therefore, with no diagonal loading, and in the ideal case of no environmental mismatch or snapshot deficiency, the peak-to-background ratio at the output of the MVDR should be 30 dB.

Since there are not enough snapshots available in our case, a small amount of loading is chosen in order to invert the cross-spectral density matrix. More precisely, $\gamma/\lambda_n = 10^{-7}$ (or −70 dB) was found sufficient to stabilize the matrix with respect to inversion. This corresponds to the first case of the bias theory for which γ is much smaller than the eigenvalues of the CSDM associated with noise; therefore we can expect the signal and noise power outputs to be shifted by $2[\gamma/\lambda_n]_{\text{in dB}} = 2 \times (-90 + 20) = -140$ dB. However, since the ambiguity surfaces are normalized to have a maximum of 0 dB prior to plotting, and both the signal and noise are biased in the same way, this 140 dB bias is not visible in the MVDR plots, as seen in the pictures obtained by Michalopoulou and Porter.[20] One can also notice that a peak-to-background ratio of about 12 dB was obtained instead of the theoretical 30 dB. Indeed, the total noise level, expressed as the sum of the noise eigenvalues, remains the same between the ideal and snapshot-deficient scenarios, while the number of noise eigenvalues decreases from LN to 10. We therefore expect the snapshot-deficient noise eigenvalues to be $10 \log[(LN)/10] = 9.8$ dB more powerful than the ideal noise eigenvalues. The inevitable environmental mismatch is responsible for the remaining discrepancy between the theoretical and experimental coherent MVDR peak-to-background ratios.

As Song also pointed out, the white noise constraint method, in contrast to the MVDR, takes advantage of this power bias.[9] Indeed the WNCM sets the white noise gain to that of the MVDR until the norm of the MVDR weight vector no longer satisfies the inequality constraint. When this happens, the CSDM is loaded causing the white noise gain to decrease and eventually reach a minimum value determined by the constraint δ itself.

Therefore, in our white noise constraint plots, the noise should be 140 dB lower than its actual physical value everywhere except around the location of the source for which the MVDR weight vector's norm becomes very large. The white noise gain at the peak is then set by the value of the constraint, and is −3.5 dB for the coherent case in Figs. 3 and 4. In other words, the peak to background ratio is expected to be given by [bias+peak-to-background ratio without loading—white noise constraint] = −140 − 12 + 3.5 ≈ −148 dB, which is consistent with the 146 dB dynamic range found experimentally.

To test the hypothesis that the bias present in our results is indeed dependent on the amount of loading, we increased the loading by steps of 2 dB in Fig. 6 and looked at the evolution of the peak-to-background ratio in decibels. The dotted lines correspond to the case previously depicted of $\gamma/\lambda_n = -70$ dB. For $\gamma/\lambda_n \in [-75, -10]$ dB, the peak-to-
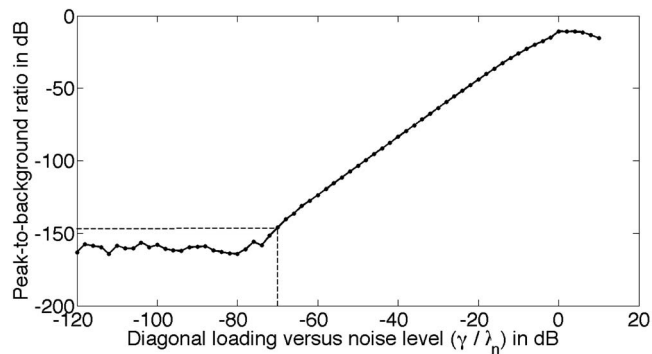
FIG. 6. Effect of the amount of diagonal loading on the bias for broadband coherent white noise processing in a snapshot-deficient scenario. As the theory suggests, the peak-to-background ratio is varying with respect to the loading with a slope of 2. The dashed lines correspond to the example previously treated.

background ratio is reduced by the diagonal loading with a slope of 2, as is found theoretically. As the loading continues to increase, the slope of the peak-to-background ratio approaches one, and eventually reaches the −12 dB ratio of the MVDR processing. When the loading gets very small in comparison to the noise eigenvalues, $\gamma/\lambda_n = -75$ dB, the matrix is increasingly ill-conditioned and the effect of loading becomes negligible. Hence the peak-to-background ratio reaches the machine precision limit of −160 dB.

As seen in Fig. 4, the MP broadband coherent white noise constraint processor localizes the source successfully while reducing the sidelobe level to −146 dB. However such a dynamic range does not imply that it can localize sources at extremely low SNR. There is in fact a threshold SNR, discussed by Baggeroer *et al.*, in the single frequency context, under which the algorithm fails to localize the source.[10] We illustrate the issue by implementing a mismatch-free simulation of the Hudson Canyon environment in a snapshot-deficient scenario for which the SNR at each receiver was reduced by steps of 2 dB. This was done by constructing ten individual realizations of signal plus white noise, and implementing the CSDM associated. Figure 7 represents the peak-
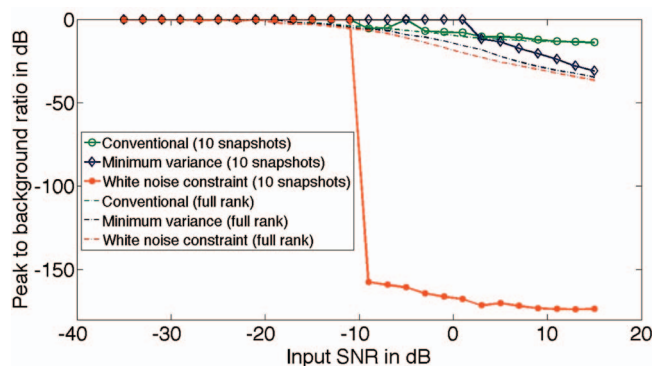
FIG. 7. Effect of the signal-to-noise ratio on the detectability of the source using the coherent broadband MP algorithm in a snapshot-deficient scenario. A peak-to-background level of 0 dB corresponds to a failure of the algorithm to localize the source. For the WNCM algorithms, a set of constraints ranging from 0.5 to 6 dB down the maximum by steps of 0.25 dB is applied, and the largest peak-to-background ratio obtained for each input SNR is represented. Despite the large dynamic range provided by the white-noise constraint, it cannot localize sources at really low SNR.
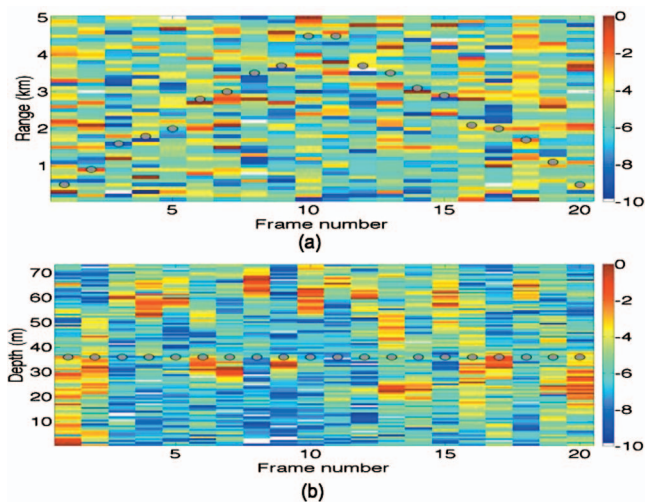
FIG. 8. Source track obtained using the incoherent minimum variance algorithm in the presence of mismatch. The black circles indicate the true source positions and the white color corresponds to a level beyond the dynamic range. (a) Range track at source depth of 36 m and (b) depth track along the estimated range track.

FIG. 9. Source track obtained using the coherent MP minimum variance algorithm in the presence of mismatch. The black circles indicate the true source positions and the white color corresponds to a level beyond the dynamic range. (a) Range track at source depth of 36 m and (b) depth track along the estimated range track.

to-background ratio obtained with the conventional, minimum variance, and white-noise constraint MP coherent broadband processor as the input SNR varies. A peak-to-background ratio of 0 dB corresponds to cases for which the algorithm failed to localize the source. One can note that despite having a 20 dB theoretical array plus processing gain, the coherent white noise constraint algorithm in this snapshot-deficient scenario does not localize a source with less than −10 dB SNR. This is due to the smaller number of eigenvalues associated with noise, discussed previously, effectively lowering the gain to 10 dB. In contrast, the curves obtained with a full rank CSDM matrix show no bias and localize sources as weak as −20 dB SNR. While the coherent white-noise constraint method displays very low sidelobes and a better performance than the other algorithms at low SNR, it nevertheless fails at a SNR lower than −10 dB despite its very large dynamic range.

The excess dynamic range associated with the broadband WNC processor is therefore convenient for localization purposes but care must be taken when making any physical interpretation.

V. SOURCE TRACKING IN RANGE AND DEPTH

In the specific example discussed in Sec. III, both MVDR and WNCM versions of the broadband coherent algorithm developed by Michalopoulou and Porter localized the source, but the white noise constraint displayed much lower sidelobes. How robust is this algorithm when the environmental mismatch is increased and the position of the source changed? To address this question, we processed each of the 20 frames present in the outgoing and incoming legs, and introduced a slight sound speed mismatch. The localization performance in range and depth of the coherent and incoherent MVDR and WNCM matched-field processors was then compared.

Figure 8(a) displays the range slices obtained with the incoherent minimum variance processor for each source po-

sition, and Fig. 8(b) shows the depth track intersecting those estimated ranges. The known source ranges and depths are indicated by the black circles. Figures 9(a) and 9(b) are the corresponding figures obtained using the coherent MP MVDR algorithm. In this case, the mismatch is sufficient to lose track of the source at most positions for both incoherent and coherent processors. The sidelobe level is, however, lower in the coherent case.
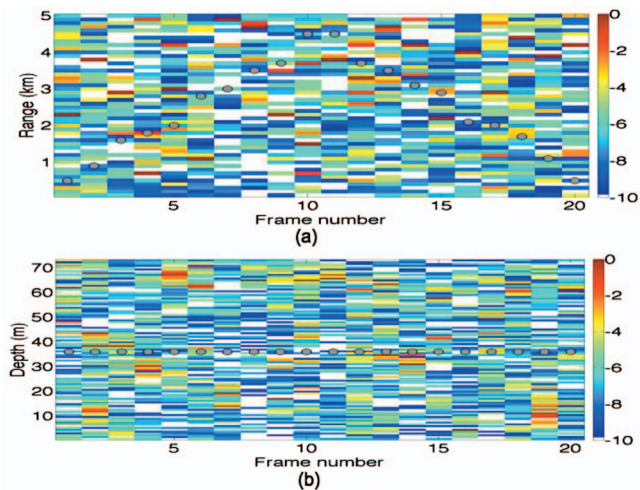
Figures 10(a), 10(b), 11(a), and 11(b), show the same type of plots using the incoherent and coherent MP white noise constraint algorithms. This time the source is tracked in range and depth in both cases, with a significantly lower sidelobe level obtained when using the coherent MP processor. The white noise constraint algorithm is therefore more robust to environmental mismatch than the MVDR and pro-
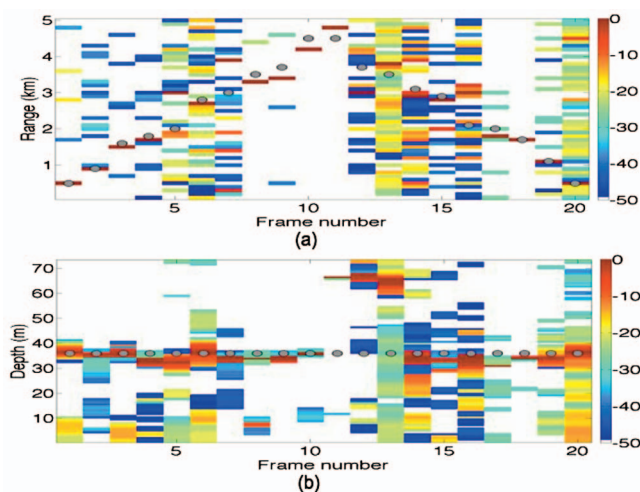
FIG. 10. Source track obtained using the incoherent white noise constraint algorithm in the presence of mismatch. The black circles indicate the true source positions and the white color corresponds to a level beyond the dynamic range. (a) Range track at source depth of 36 m and (b) depth track along the estimated range track.
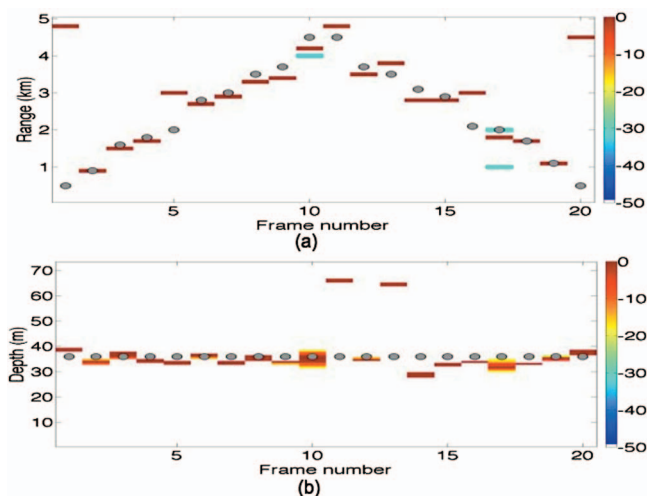
FIG. 11. Source track obtained using the coherent MP white noise constraint algorithm in the presence of mismatch. The black circles indicate the true source positions and the white color corresponds to a level beyond the dynamic range. (a) Range track at source depth of 36 m and (b) depth track along the estimated range track.

duces low sidelobes when combined with a coherent broadband algorithm like the one developed by Michalopoulou and Porter.

VI. SUMMARY AND CONCLUSION

The problem of broadband acoustic source localization in shallow water has been addressed. As previously shown, exploiting the information across frequency is essential to perform a successful localization in this type of environment. Two approaches were compared, the first consists in incoherently averaging the information across frequencies, and the second, developed by Michalopoulou and Porter, treats the multiple frequencies information coherently. Applying the very common MVDR estimator to both methods revealed that the MFP discrimination is indeed improved by the use of a coherent algorithm, but, on the other hand, that sensitivity to mismatch is still problematic.

To address this issue, the robust white noise constraint estimator has been applied to the coherent and incoherent approaches. Both methods were able to track a source accurately and continuously using experimental data, in the presence of enough mismatch to lead to erroneous MVDR detections. The coherent white noise constraint processor has also been shown through simulation to perform better than the conventional and MVDR estimators at low SNR. Application of the WNCM algorithm to the MP broadband coherent method results in a robust, discriminating, high resolution processor well suited for practical applications.

ACKNOWLEDGMENTS

[1] A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, "An overview of matched fields methods in ocean acoustics," IEEE J. Ocean. Eng. **18**, 401–424 (1993).

[2] A. Tolstoy, *Matched Field Processing for Underwater Acoustics* (World Scientific, Singapore, 1992).

[3] J. N. Maksym, "A robust formulation of an optimum cross-spectral beamformer for line arrays," J. Acoust. Soc. Am. **65**, 971–975 (1979).

[4] H. S. Schmidt, A. B. Baggeroer, W. A. Kuperman, and E. K. Scheer, "Environmentally tolerant beamforming for high-resolution matched field processing: Deterministic mismatch," J. Acoust. Soc. Am. **73**, 813–825 (1979).

[5] J. L. Krolik, "Matched-field minimum variance beamforming in a random ocean channel," J. Acoust. Soc. Am. **92**, 1408–1419 (1992).

[6] J. V. Candy and J. Sullivan, "Model-based identification: An adaptive approach to ocean-acoustic processing," IEEE J. Ocean. Eng. **21**, 273–289 (1996).

[7] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," IEEE Trans. Acoust., Speech, Signal Process. **35**, 1365–1376 (1987).

[8] A. B. Baggeroer and H. Cox, "Passive sonar limits upon nulling multiple moving ships with large aperture arrays," *Proceedings of the IEEE, 33rd Asilomar Conference on Signals, Systems Computing*, Pacific Grove, CA, 1999, pp. 103–108.

[9] H. Song, W. A. Kuperman, W. S. Hodgkiss, P. Gerstoft, and J. S. Kim, "Null broadening with snapshot-deficient covariance matrices in passive sonar," IEEE J. Ocean. Eng. **28**, 250–261 (2003).

[10] A. B. Baggeroer, W. A. Kuperman, and H. Schmidt, "Matched-field processing: Source localization in correlated noise as an optimum parameter estimation problem," J. Acoust. Soc. Am. **83**, 571–587 (1988).

[11] C. S. Clay, "Optimum time doman signal transmission and source location in a waveguide," J. Acoust. Soc. Am. **81**, 660–664 (1987).

[12] S. Li and C. D. Clay, "Optimum time domain signal transmission and source location in a waveguide: Experiments in an ideal wedge waveguide," J. Acoust. Soc. Am. **82**, 1409–1417 (1987).

[13] C. S. Clay and S. Li, "Time domain signal transmission and source location in a waveguide: Matched filter and deconvolution experiments," J. Acoust. Soc. Am. **83**, 1377–1383 (1988).

[14] R. K. Brienzo and W. S. Hodgkiss, "Broadband matched-field processing," J. Acoust. Soc. Am. **94**, 2821–2831 (1993).

[15] L. N. Frazer and P. I. Pecholcs, "Single-hydrophone localization," J. Acoust. Soc. Am. **88**, 995–1002 (1990).

[16] P. Hursky, M. B. Porter, and M. Siderius, "High-frequency (8-16 kHz) model-base source localization," J. Acoust. Soc. Am. **115**, 3021–3032 (2004).

[17] E. K. Westwood,, "Broadband matched-field source localization," J. Acoust. Soc. Am. **91**, 2777–2789 (1992).

[18] D. P. Knobles and S. K. Mitchell, "Broadband localization by matched fields in range and bearing in shallow water," J. Acoust. Soc. Am. **96**, 1813–1820 (1994).

[19] S. P. Czenszak and J. L. Krolik, "Robust wideband matched-field processing with a short vertical array," J. Acoust. Soc. Am. **101**, 749–759 (1997).

[20] Z.-H. Michalopoulou and M. B. Porter, "Matched-field processing for broad-band source localization," IEEE J. Ocean. Eng. **21**, 384–392 (1996).

[21] Z.-H. Michalopoulou, "Robust muli-tonal matched-field inversion: A coherent approach," J. Acoust. Soc. Am. **104**, 163–170 (1998).

[22] G. J. Orris, M. Nicholas and J. S. Perkins,, "The matched-phase coherent multi-frequency matched-field processor," J. Acoust. Soc. Am. **107**, 2563–2575 (2000).

[23] H. L. Van Trees, *Optimum Array Processing, Part IV* "Detection, estimation, and modulation theory," (Wiley-interscience, New York, 2002).

[24] H. A. Chandler, C. Feuillade, and G. B. Smith, "Sector-focused processing for stabilized resolution of multiple acoustic sources," J. Acoust. Soc. Am. **97**, 2159–2172 (1995).

[25] W. M. Carey, J. Doutt, R. B. Evans, and L. M. Dillman, "Shallow-water sound transmission measurements on the New Jersey Continental Shelf," IEEE J. Ocean. Eng. **20**, 321–335 (1995).

[26] M. Porter, "The Kraken normal mode program (sm245)," SACLANT Undersea Research Center, La Spezia, Italy, 1991.

Comparison of simulations and data from a seismo-acoustic tank experiment

Jon M. Collis[a) and William L. Siegmann
Rensselaer Polytechnic Institute, Troy, New York 12180

Michael D. Collins, Harry J. Simpson, and Raymond J. Soukup
Naval Research Laboratory, Washington, D.C. 20375

A tank experiment was carried out to investigate underwater sound propagation over an elastic bottom in flat and sloping configurations. The purpose of the experiment was to evaluate range-dependent propagation models with high-quality experimental data. The sea floor was modeled as an elastic medium by a polyvinyl chloride slab. The relatively high rigidity of the slab requires accounting for shear waves in this environment. Acoustic measurements were obtained along virtual arrays in the water column using a robotic apparatus. Elastic parabolic equation solutions are in excellent agreement with data. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2756968]

Pages: 1987–1993

I. INTRODUCTION

The ocean bottom has a major effect on shallow water sound propagation. The sediment can often be modeled as a fluid, but shear effects are important in some cases. The interface between the ocean and sediment can be modeled as horizontal in some cases, but its slope must be taken into account in many applications. Previous calculations have combined bottom effects and sloping fluid-solid interfaces, using parabolic equation,[1,2] spectral,[3–5] and hybrid[6,7] approaches. However, the combination of shear effects and bottom slope is a difficult problem to model accurately and efficiently. The parabolic equation method is an effective approach for solving such range-dependent seismo-acoustics problems,[8,9] including anisotropic[10] and interface wave effects.[11] Recent progress has improved the accuracy of this approach for problems involving sloping interfaces.[12,13]

Due to the difficulty in collecting high-quality experimental data, there have been relatively few experiments conducted which were capable of verifying the predictions of underwater sound models. Recent experiments focused on propagation up a wedge,[14] three-dimensional effects,[15] and rough surface backscattering.[16] Although tank experiments of acoustic propagation have established the significance of elastic effects they did not involve detailed field comparisons.[17–19] As a result, no definitive reference exists which validates range-dependent elastic parabolic equation models against high-quality experimental data.

In this paper, we compare parabolic equation solutions with data from an experiment involving a slab of polyvinyl chloride (PVC) that was suspended in a tank of water. For this scaled model of both flat and sloping elastic ocean bottoms, high-quality data were obtained by using a robotic apparatus to move a hydrophone along virtual arrays. The elas-

tic parameters of the slab were measured prior to the experiment. We were unable to obtain acoustic/fluid parabolic equation solutions that agree even approximately with the data. When shear was included in the model, we were able to achieve excellent agreement by fine tuning the geometric parameters. The experiment is reviewed in Sec. II. A discussion of the computer scale model and inversion procedures is given in Sec. III. Comparisons with parabolic equation solutions are presented in Sec. IV.

II. TANK EXPERIMENT

A series of laboratory experiments with scale models of an elastic ocean bottom are being performed in order to test predictions of seismo-acoustic models. Here we report on the first Elastic Parabolic Equation Experiment (EPE 1), which was carried out in April 2004 inside a large fresh water tank at the Naval Research Laboratory.[20] An elastic bottom was modeled using a PVC slab ($122 \times 122 \times 10$ cm) from San Diego Plastics. The slab was suspended in water by cables, which were attached to its corners to minimize reflections. Source and receiver hydrophones were positioned over the slab using a robotic apparatus that allowed for accurate positioning. The slab, cables, source and receiver alignments, and robotic arms are illustrated in Fig. 1. The source, closer to the edge of slab, was fixed while the receiver moved at 2 mm increments away from the source. (An experiment with similar instrumentation and a milled PVC slab was performed a month later and is documented in Ref. 21.)

During EPE-1 the water temperature was maintained so that the sound speed remained within 1 m/s of 1482 m/s. As discussed in Ref. 20, estimates for the elastic properties of the slab in the band 300 kHz–1.5 MHz were obtained by transmitting pulses through a sample of the material. Results at 300 kHz are summarized in Table I, with estimated relative errors shown in the second column. Since this frequency band is at the high end of the transmitted frequency band of

[a)Present address: Boston University, Boston, Massachusetts. Electronic mail:jcollis@bu.edu
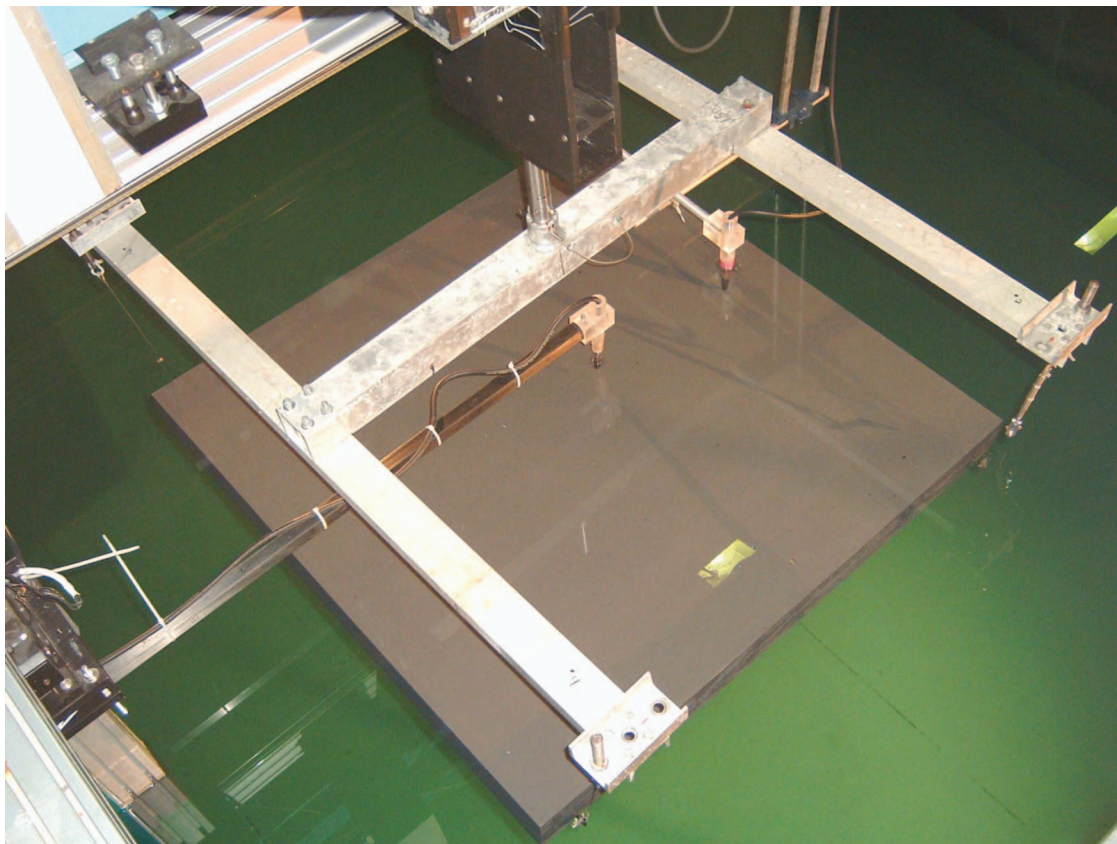
FIG. 1. The experimental setting; the PVC slab is suspended in water by cables attached near its corners. The robotic apparatus which positions the source and receiver hydrophones is shown.

the acoustic tank experiment, using the values at lower frequencies may introduce an extrapolation error.

The transmitted waveform was an impulse with a flat frequency spectrum over the band 100–300 kHz. A reference measurement was made by positioning the source and receiver 1 m apart and measuring the pressure produced from a shaped pulse which was flat within the source band. Figure 2 shows the temporal and frequency responses for the reference measurement. Measurements were recorded at 8192 points with a 0.5 μs sampling interval. Windowing was applied to eliminate reflections from the hardware and the walls of the tank. The windowed time series were Fourier transformed, resulting in 4096 data points which spanned 1 MHz in the frequency domain. A transfer function was obtained by multiplying the frequency response with the reciprocal of the spectrum of the reference pulse.

Experimental runs were conducted in two environments with different source and receiver configurations. To show

TABLE I. Estimated physical values, with associated error estimates, of elastic PVC properties at 300 kHz (see Ref. 20).

Parameter	Value	Relative Error ($\pm\%$)	Absolute Error
Density (kg/m^3)	1378	0.5	7
Compressional speed (m/s)	2290	0.5	10
Compressional attenuation (dB/m)	0.33	2	0.01
Shear speed (m/s)	1050	0.5	5
Shear attenuation (dB/m)	1.00	4	0.04

the analogy between propagation in the tank and an ocean waveguide, we present acoustic propagation calculations at a scale of 1000:1 in Fig. 3, where the frequencies and lengths have been appropriately modified. The (constant) compressional and shear attenuation α_p and α_s in the units dB/λ (λ being the acoustic wavelength) are invariant under the scaling, because they are implicitly scaled with the wavelength. The first environment is essentially range independent (nearly horizontal slab), and a sample of calculated transmission loss contours is shown in Fig. 3(a) for a 100 Hz source. Also shown are the geometrical parameters: source depth z_s (which can be near the middle of the water column or near the bottom), depth of the slab below the source z_0, and depth of the opposite edge of the slab z_1. The receiver depth z_r, is not shown. Nominal values of these parameters are given in the second column of Table II. For the first two runs, the slab was suspended nearly parallel to the air-water interface. The propagation track was centered on the slab, with the source 15 cm away from the edge. The receiver was moved horizontally between 25 and 135 cm from the edge of the slab nearest the source to produce a virtual horizontal array with a spacing of 2 mm. The slab attenuated acoustic energy sufficiently to prevent spurious reflections from the slab bottom. For the last two runs, the environment is range dependent, with the slab sloping as illustrated in Fig. 3(b). Nominal values of the geometrical parameters shown in Fig. 3(b) are given in the second column of Table III. In the time period between the positioning of the source and receiver in the tank and the data collection, it is estimated that the depth of
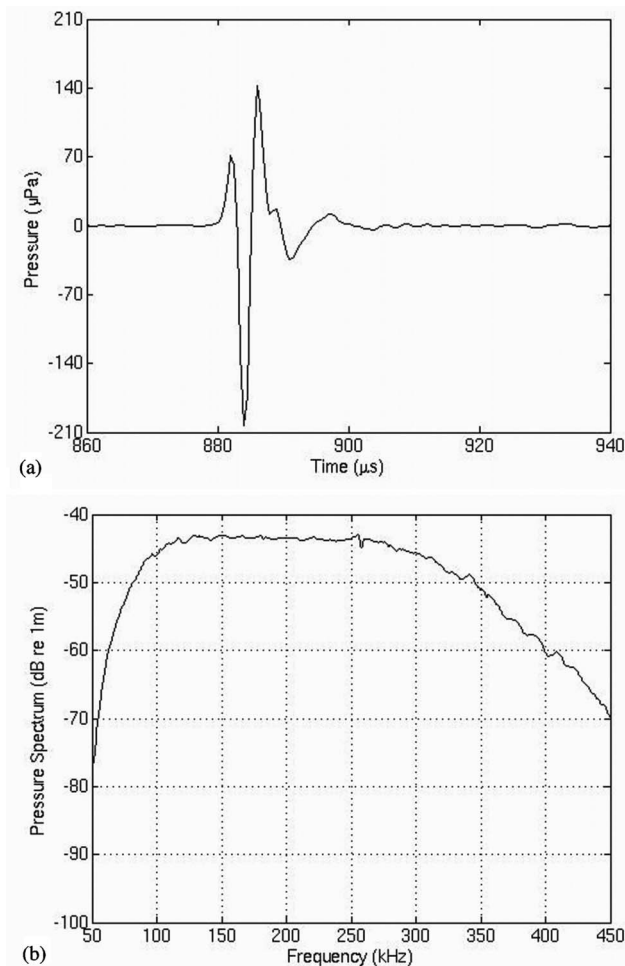
FIG. 2. Reference measurements of the source and receiver separated by 1 m. (a) Temporal response plot of pressure (μPa) vs time (μs). (b) The frequency response plotted as power spectral density vs frequency (kHz). The frequency response was designed to be flat in the 100–300 kHz band.

FIG. 3. Calculated transmission loss contours for the compressional field of a 100 Hz mid-depth source using scaled parameters for length and frequency: (a) flat case, (b) sloped case.

the slab, source and receiver could have decreased by as much as 6 mm due to evaporation of water from the tank. This evaporation affected the geometrical parameters in an analogous manner for both the flat and sloped slab cases, as the source and receiver were re-positioned prior to the sloped slab data acquisition. Additional uncertainty in the geometrical parameters could arise from a non-central acoustic source location inside the spherical object (of radius 1 cm) and slight deviations from flatness for the slab.

III. PROPAGATION CALCULATIONS

Experimental data were compared with calculations from a parabolic equation model suitable for range-dependent environments and fluid or elastic media. Because parabolic equation solutions are exact in range-independent environments and highly accurate in range-dependent environments, we anticipate close comparisons with data, provided accurate estimates of geometric and material parameter values are available. Initial comparisons of data and calculations using the measured and recorded parameter values suggested the need for refinement. Two inversion procedures, using an elastic parabolic equation model, were used to determine precise estimates. First, forward iterations at selected

frequencies were performed, and parameter values were found that minimized the vector norm of transmission loss differences. Second, simulated annealing inversions[22–24] were run at those frequencies to refine the values. Deep fades (local minima) in transmission loss patterns were sufficiently rare and narrow enough to not affect the inversion results. The simulated annealing algorithm is based on the coordinate rotation approach of Ref. 22. We selected parameter values by averaging results over the five nominal frequencies 100, 150, 200, 250, and 300 kHz.

Results from the two inversion processes are given in Tables II and III for the flat and sloped cases. When comparing values obtained from the inversions with the nominally recorded values, the differences are generally less than 10%. The maximum difference of 22% occurs with the shallow receiver in the sloped case, although the absolute error for this case is small. The largest absolute error (12 mm) occurs for the deep source in the sloping configuration and probably results from a combination of uncertainties due to evaporation and in slab/source depths. The actual differences for all other values, of 1 cm or less, reflect the precision required for accurate comparisons. For example, the flat case is nominally horizontal within 1 mm, but this deviation is sufficient to distort the comparisons observably. The refinements in values obtained by using simulated annealing after forward iterations are small, but they result in improvements in all cases.

TABLE II. Inversion results for geometric parameters: for the flat slab case. Nominal values are according to the experimental plan; other values are from two inversion procedures. Estimates from simulated annealing yield better agreement with data. Errors are between nominal and simulated annealing results.

Flat case	Nominal	Iterative	Simulated annealing	Absolute error	Relative error ($\pm\%$)
z_s mid (cm)	7.50	6.91	6.91	0.59	7.9
z_s deep (cm)	14.55	13.63	13.68	0.87	6.0
z_r (cm)	14.55	13.74	13.71	0.84	5.8
z_0 (cm)	15.00	14.44	14.47	0.53	3.6
z_1 (cm)	15.00	14.55	14.54	0.46	3.1

IV. COMPARISONS OF MEASUREMENTS AND CALCULATIONS

Data comparisons were made with parabolic equation implementations for fluid[25] and elastic media. The elastic model uses a variable rotated elastic parabolic equation.[13] Results are presented for selected geometries and source frequencies of 125,200, and 275 Hz. For all calculations, the water sound speed is 1482 m/s and the slab density, sound speeds, and attenuations are the measured values from Table I. Computations were made to the effective maximum range of 120 cm.

Range-independent propagation

We first consider examples of the nominally flat case, which actually has weak range dependence. The source is located at either mid-depth ($z_s=6.91$ cm) or deep ($z_s=13.68$ cm). The values of z_r, z_0, and z_1 are given in the fourth column of Table II, where the depth z_1 is measured at $r=106.8$ cm. Figure 4 shows comparisons between the elastic solution and data for the mid-depth source. There is excellent agreement between these calculations and the data, with only subtle differences observed in the patterns. The elastic parabolic equation solution closely tracks detailed variations in the transmission loss. Deep source results are illustrated in Figs. 5 and 6 for both fluid and elastic parabolic equation calculations. Solutions in Fig. 5 for the fluid parabolic equation show significant differences from data and only manage to capture the basic trend of the data curves. Close to the source, the fluid solutions track the interference pattern and agree fairly well with measurements, but they seriously underestimate the loss as range increases. In contrast, solutions from the elastic parabolic equation are plotted in Fig. 6. For all three frequencies these solutions provide close matches to both the interference pattern and amplitude

of the measured transmission loss. Although the fields differ dramatically for the two source positions, both are accurately modeled by the elastic parabolic equation solutions. The field structure is somewhat more complicated with the deep source position, but the solutions still handle the precise phase and amplitude changes of the pattern well. Beyond the edge of the slab, the field differs slightly from the data.

Range-dependent propagation

Figure 7 shows elastic model calculations and data for range-dependent propagation over an approximately 4.8° slope, for the case where the source is at mid-depth and the receiver is near-surface. The values of z_s, z_r, z_0, and z_1 used in the calculations are in the fourth column of Table III, where the depth at z_1 is measured at $r=106.5$ cm. There is excellent agreement between the calculated and recorded values, with fine details of the field resolved to high accuracy. An extensive collection of additional sloping bottom cases for other frequencies and source and receiver positions was investigated, and the fields in all cases are very well represented by the elastic parabolic equation calculations.

V. CONCLUSIONS

Our main conclusion is the demonstration of excellent agreement between elastic parabolic equation solutions and experimentally recorded values for sound propagation in water overlying an elastic slab. The agreement confirms that the elastic parabolic equation is capable of resolving both nearly range-independent and range-dependent problems efficiently and precisely. The results obtained for the range-dependent sloped case provide support for the validity of the parabolic equation derivation in its elastic formulation. The fine details of the transmission loss pattern that are resolved by the elastic solutions represent benchmark accuracy agreement with

TABLE III. Inversion results for geometric parameters: sloped case. Nominal values are according to the experimental plan; other values are from two inversion procedures. Estimates from simulated annealing yield better agreement with data. Errors are between nominal and simulated annealing results.

Sloped case	Nominal	Iterative	Simulated annealing	Absolute error	Relative error ($\pm\%$)
z_s mid (cm)	6.90	6.34	6.34	0.56	8.2
z_s deep (cm)	13.55	12.43	12.35	1.2	8.9
z_r (cm)	2.00	1.57	1.56	0.44	22
z_0 (cm)	13.80	13.34	13.29	0.51	3.7
z_1 (cm)	4.95	4.55	4.54	0.41	8.2
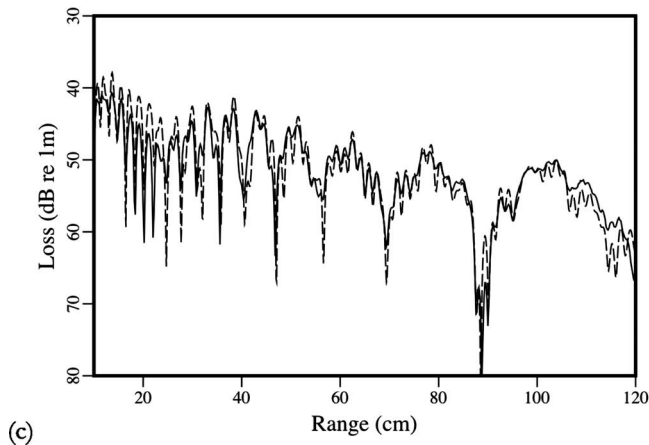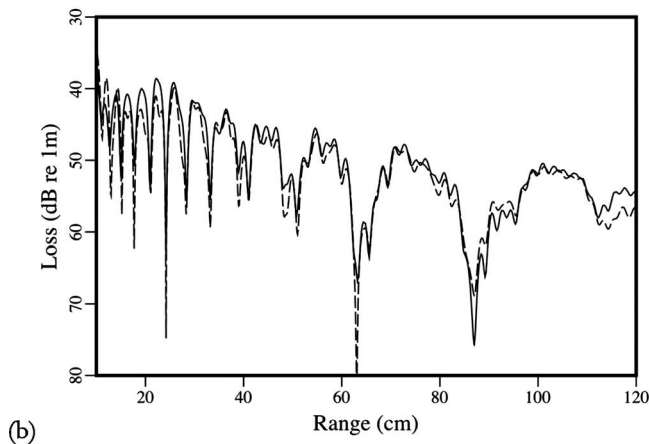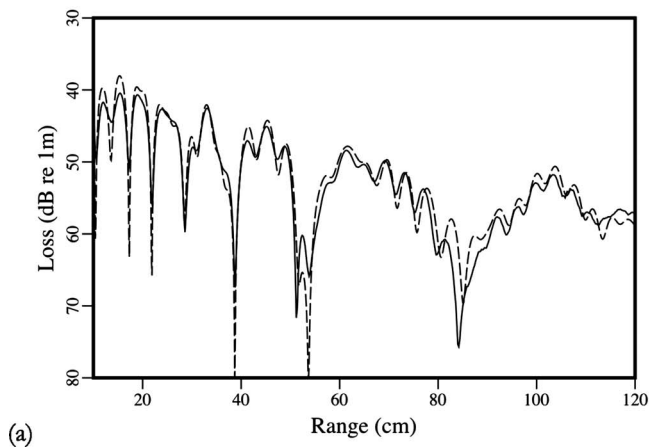
FIG. 4. Transmission loss vs range for the flat case, mid-depth source at 6.91 cm and near-bottom receiver at 13.71 cm. Comparisons show data (solid curve) and calculations from the elastic parabolic equation (dashed curve), for source frequencies: (a) 125 kHz, (b) 200 kHz, and (c) 275 kHz.

the data. Another conclusion is the demonstration of the expected poor agreement between fluid parabolic equation solutions and recorded values. In combination with the excellent agreement of the elastic solution, this result shows that the fluid parabolic equation is unsatisfactory for modeling underwater sound propagation over bottoms with significant elasticity.

Iteration by inversion methods for values of the geometric parameters was essential as the evaporation of water from the tank and other uncertainties in the source and receiver
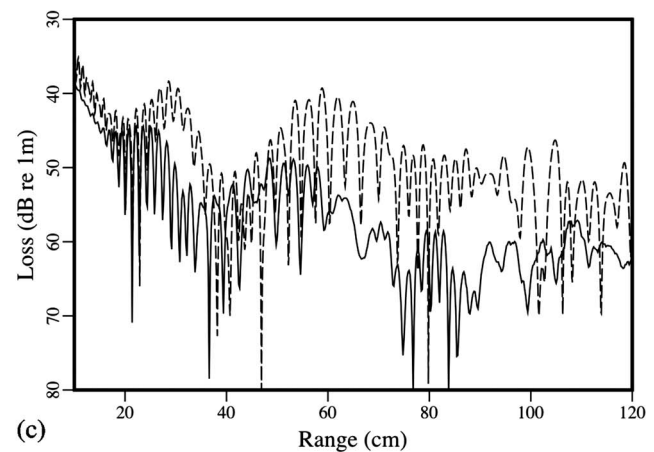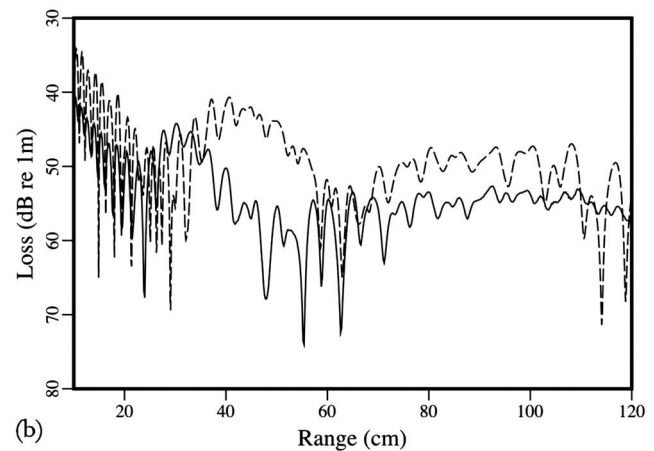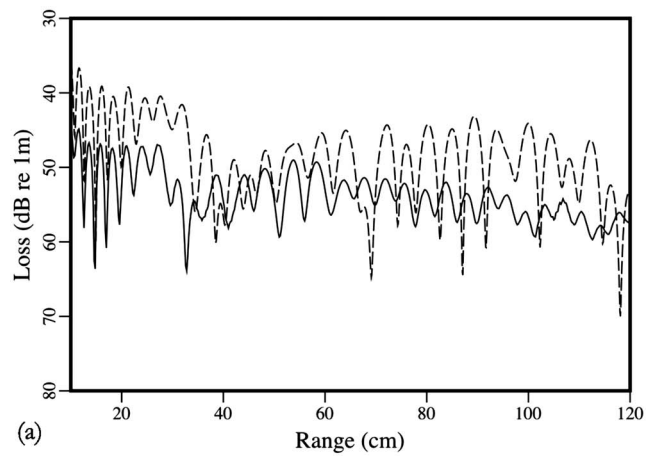
FIG. 5. Transmission loss vs range for the flat case, deep source at 13.69 cm, and near-bottom receiver at 13.71 cm. Comparisons show data (solid curve) and calculations from the fluid parabolic equation RAM (dashed curve), for source frequencies: (a) 125 kHz, (b) 200 kHz, and (c) 275 kHz.

position and the slab surface hat to be taken into account. The success of the elastic PE model in obtaining excellent agreement with data, given these adjustments, illustrates the sensitivity of acoustic propagation in a simple waveguide to small changes in the experimental geometry. The detailed agreement illustrates that the elastic PE code could be used to successfully model a simulated waveguide for which the
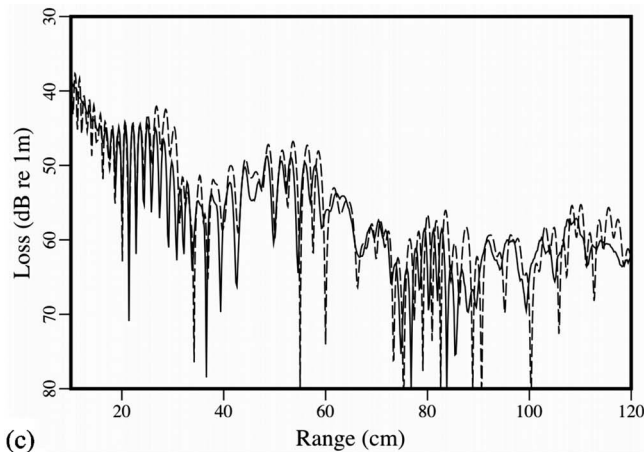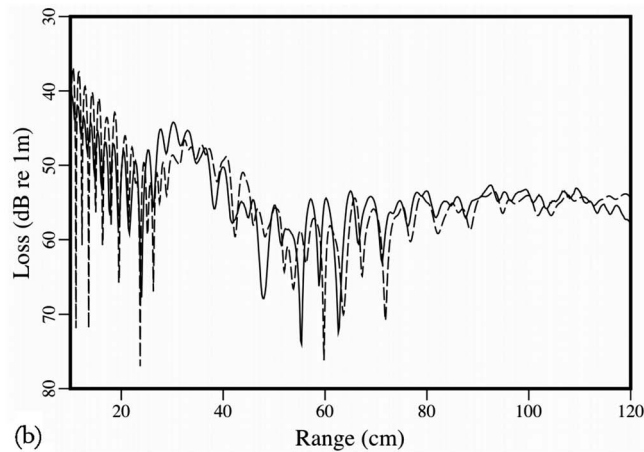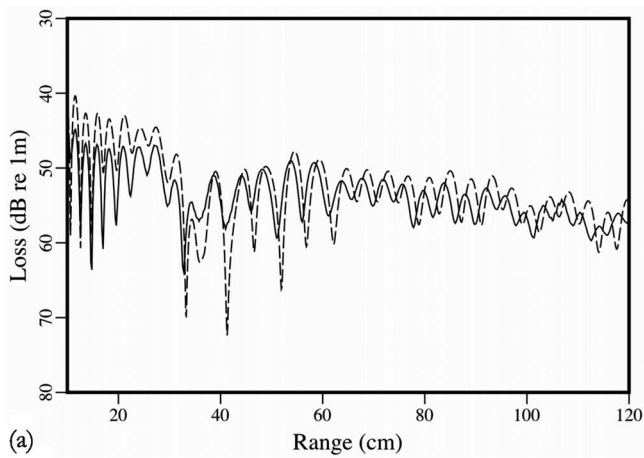
FIG. 6. Transmission loss vs range for the flat case, deep source at 13.69 cm and near-bottom receiver at 13.71 cm. Comparisons show data (solid curve) and calculations from the elastic parabolic equation (dashed curve), for source frequencies: (a) 125 kHz, (b) 200 kHz, and (c) 275 kHz.

FIG. 7. Transmission loss vs range for the sloped case, mid-depth source at 6.34 cm and near-surface receiver at 1.56 cm. Comparisons show data (solid curve) and calculations from the elastic parabolic equation (dashed curve), for source frequencies: (a) 125 kHz, (b) 200 kHz, and (c) 275 kHz.

shere properties of the bottom are of great importance. Future work will focus on more complex varieties of range dependence.

ACKNOWLEDGMENTS

[1] M. D. Collins, "A higher-order parabolic equation for wave propagation in an ocean overlying an elastic bottom," J. Acoust. Soc. Am. **86**, 1459–1464 (1989).

[2] M. D. Collins, "The rotated parabolic equation and sloping ocean bottoms," J. Acoust. Soc. Am. **87**, 1035–1037 (1990).

[3] J. T. Goh, H. Schmidt, P. Gerstoft, and W. Seong, "Benchmarks for validating range-dependent seismo-acoustic propagation codes," IEEE J. Ocean. Eng. **22**, 226–236 (1997).

[4] I. T. Lu and L. B. Felsen, "Adiabatic transforms for spectral analysis and synthesis of weakly range-dependent shallow ocean Green's functions," J. Acoust. Soc. Am. **81**, 897–911 (1987).

[5] J. T. Goh and H. Schmidt, "Validity of spectral theories for weakly range-dependent ocean environments—numerical results," J. Acoust. Soc. Am. **95**, 727–732 (1994).

[6] H. Schmidt, W. Seong, and J. T. Goh, "Spectral super-element approach to range-dependent ocean acoustic modeling," J. Acoust. Soc. Am. **98**, 465–472 (1995).

[7] J. T. Goh and H. Schmidt, "A hybrid coupled wavenumber integration approach to range-dependent ocean acoustic modeling," J. Acoust. Soc. Am. **100**, 1409–1420 (1996).

[8] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (Springer-Verlag, New York, 2000), pp. 325–389.

[9] M. D. Collins, "Higher-order parabolic approximations for accurate and stable elastic parabolic equations with application to interface wave propagation," J. Acoust. Soc. Am. **89**, 1050–1057 (1991).

[10] A. J. Fredricks, W. L. Siegmann, and M. D. Collins, "A parabolic equation for anisotropic elastic media," Wave Motion **31**, 139–146 (2000).

[11] W. Jerzak, W. L. Siegmann, and M. D. Collins, "Modeling Rayleigh and Stoneley waves and other interface and boundary effects with the parabolic equation," J. Acoust. Soc. Am. **117**, 3497–3503 (2005).

[12] M. D. Collins and D. K. Dacol, "A mapping approach for handling sloping interfaces," J. Acoust. Soc. Am. **107**, 1937–1942 (2000).

[13] D. A. Outing, W. L. Siegmann, M. D. Collins, and E. K. Westwood, "Generalization of the rotated parabolic equation to variable slopes," J. Acoust. Soc. Am. **120**, 3534–3538 (2006).

[14] S. A. L. Glegg, G. B. Deane, and I. G., House, "Comparison between theory and model scale measurements of three-dimensional sound propagation in a shear supporting penetrable wedge," J. Acoust. Soc. Am. **94**, 2334–2342 (1993).

[15] J. P. Sessarego, "Scaled models for underwater acoustics and geotechnics applications," *Proceedings of the Sixth European Conference on Underwater Acoustics*, Gdansk, Poland (2002).

[16] R. J. Soukup, G. Canepa, H. J. Simpson, J. E. Summers, and R. F. Gragg, "Small-slope simulation of acoustic backscatter from a physical model, J. Acoust. Soc. Am., in press.

[17] I. Tolstoy, "Shallow water test of the theory of layered wave guides," J. Acoust. Soc. Am. **30**, 348–361 (1958).

[18] I. Tolstoy, "Guided waves in a fluid with continuously variable velocity overlying an elastic solid:Theory and experiment," J. Acoust. Soc. Am. **32**, 81–87 (1960).

[19] R. K. Eby, A. O. Williams, R. P. Ryan, and P. Tamarkin, "Study of acoustic propagation in a two-layered model," J. Acoust. Soc. Am. **32**, 88–99 (1960).

[20] R. J. Soukup, H. J. Simpson, E. C. Porse, J. E. Summers, and R. F. Gragg, "Geoacoustic physical modeling elastic parabolic equation 1 (GPM EPE1) experiment: Measurement report and acoustic data," U.S. Naval Research Laboratory Memo. Rep. 7140-04-8826 (2004).

[21] R. J. Soukup, H. J., Simpson, E. C. Porse, J. E. Summers, and R. F. Gragg, "Geoacoustic physical modeling rough surface scattering experiment 2 (GPM RSS2): Measurement report, acoustic data and profilometry," U.S. Naval Research Laboratory Memo. Rep. 7140-04-8827 (2004).

[22] M. D. Collins and L. Fishman, "Efficient navigation of parameter landscapes," J. Acoust. Soc. Am. **98**, 1637–1644 (1995).

[23] M. D. Collins, W. A. Kuperman, and H. Schmidt, "Nonlinear inversion for ocean-bottom properties," J. Acoust. Soc. Am. **92**, 2770–2783 (1992).

[24] R. J. Cederberg and M. D. Collins, "Applications of an improved self-starter to geoacoustic inversion," IEEE J. Ocean. Eng. **22**, 102–109 (1997).

[25] M. D. Collins, "User's Guide for RAM Versions 1.0 and 1.0p", ftp://ftp.ccs.nrl.navy.mil/pub/ram/RAM(2006).

Alpha-plutonium's polycrystalline elastic moduli over its full temperature range

A. Migliori,[a)] C. Pantea, H. Ledbetter, I. Stroe, J. B. Betts, J. N. Mitchell, M. Ramos, F. Freibert, D. Dooley, S. Harrington, and C. H. Mielke

Los Alamos National Laboratory, Los Alamos, New Mexico 87545

α-plutonium's volume-corrected polycrystal elastic moduli were measured between 18 K and the upper limit of its occurrence, near 400 K. The two independent moduli for a polycrystal—bulk and shear—behave smoothly, indicating no phase transition. Both moduli show the same 50% increase on cooling, an order of magnitude larger than in other metals. The Debye temperature obtained from low-temperature elastic moduli, 207 K, significantly exceeds most previous estimates. The Gruneisen parameter $\gamma = 5.3$, obtained from the temperature dependence of the bulk modulus, is intermediate among previous estimates using other approaches, α-plutonium's Poisson ratio ν is low: 0.18, nearly temperature independent, and its small decrease on warming opposes usual behavior. The high γ, large but equal bulk modulus and shear modulus fractional stiffening on cooling, and near-temperature-invariant ν are attributed to a single mechanism: 5-f electron localization—delocalization. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2767419]

I. INTRODUCTION

Stable between 0 and 404 K, plutonium's α phase, the lowest-temperature allotrope among six, shows a monoclinic crystal structure with a 16-atom unit cell ($mP16$, $P2_1/m$, C_{2h}^2, No. 11, with only three symmetry elements beyond the identity operator!). Thus, unlike the b.c.c., f.c.c., and c.p.h. crystal-structure materials that received much study of their physical properties, which often depend strongly on the crystal structure, monoclinic metals by their scarcity remain nearly nonstudied. Plutonium's atomic electronic structure, [Rn]$5f^6 7s^2$, provides no clue to its monoclinicity. The structure is highly nonsymmetrical, containing only 4 symmetry elements, compared with 48 in its sister f.c.c. δ-phase allotrope. Indeed, plutonium represents the least-symmetrical metallic element. Why does plutonium show a monoclinic space group, almost nonexistent among metals? The extensive ($\sim 300\,000$) compound literature shows that most (54%) compounds crystallize monoclinically,[1] perhaps because a monoclinic space group permits close packing, different "near-neighbor" bond lengths, and preserves interatomic-bond angles. These factors suggest that plutonium's interatomic bonding departs from purely metallic, is partly covalent, that some bonding electrons are not itinerant, but semilocalized between atoms, and that plutonium (an element) must share some bonding characteristics of compounds. Physical properties such as plutonium's high electrical resistivity and low melting point support the covalent-bonding concept. Based on short near-neighbor interatomic-bond lengths, Zachariasen and Ellinger proposed plutonium covalency about 45 years ago.[2] Covalency should appear in the elastic moduli.[3]

Here we report α-plutonium's polycrystal elastic moduli, including longitudinal modulus C_L, shear modulus G, the two independent moduli most directly obtained from the measurements, as well as others computed from these: bulk modulus B, and Young modulus E. Also, we consider the Poisson ratio ν, a dimensionless elastic-modulus ratio. From the moduli extrapolated to zero temperature, we computed the Debye temperature Θ_D. By three alternative approaches, we estimated the quintessential anharmonic property: the Gruneisen parameter γ.

Elastic moduli, their temperature, and pressure derivatives, and derived thermodynamic properties play key roles in a material's equation of state.[4,5] Much current research now proceeds on plutonium's equation of state,[6] as well as on actinide equations of state in general.[7] Because the elastic moduli determine the harmonic parameter, the Debye temperature, they relate directly to, or correlate with, an enormous range of solid-state properties.[8,9]

One can measure elastic moduli accurately and with extreme (part per million) precision. They are, therefore, often the most sensitive test for phase transformations: electronic, magnetic, structural.[10] As described in a recent review,[11] α-plutonium's low-temperature magnetic state remains controversial. Low-temperature elastic moduli, especially the bulk modulus, play a key role in confirming various theories, many being proposed recently.[12]

To measure the moduli, we used resonant-ultrasound spectroscopy (RUS).[13–15] This approach provides four principal advantages: A single set of measurements yields all independent elastic moduli, free-surface boundary conditions are easily maintained to the limit of the sensitivity of the measurement, it provides high accuracy without corrections (typically less than 1% error), and it exhibits part-per-million precision.

---

a)Electronic mail: migliori@cybermesa.com

II. MEASUREMENTS

A. Materials, thermal expansion, temperature, and mass density

Measurements were made on two batches of electrorefined ^{239}Pu. Theses batches were determined to be 99.96 wt % Pu, 115 ppm W, 49 ppm Np, 50 ppm O, 53 ppm Si, 32 ppm Am, with the sum of remaining impurities less than 25 ppm. The rectangular parallepiped resonators used for RUS and thermal-expansion measurements made above 300 K were cut from larger buttons that were arc-melted and quenched on a copper hearth, cut and polished, examined optically for voids and metallurgical defects, and remelted until metallurgical imperfections were negligible. The specimen for RUS measurements below 300 K was only arc-melted once, and was not examined metallographically. By usual metallographic cut-grind-polish methods we prepared rectangular-parallelepiped specimens for RUS measuring 0.475 cm$\times$0.377 cm$\times$0.484 cm (below 300 K), 0.265 cm$\times$0.268 cm$\times$0.270 cm (above 300 K), and approximately 0.1 cm$\times$0.3 cm$\times$0.8 cm for thermal-expansion where the measurement was conducted along the longest dimension.

Specimen density was determined at near-ambient temperature by a custom-designed immersion density apparatus using 3M Flourinert FC-43 as an immersion fluid. Temperatures were determined by platinum thermometry. Temperature errors were ±0.01 K and weighing errors were ±0.01 mg. We corrected the density at ambient temperature for thermal expansion of the immersion fluid and specimen self-heating effects. The mass density was 19.55 g/cm^3±0.02% for the RUS specimen used in measurements above 300 K and for the thermal-expansion studies. Density was 19.59 g/cm^3±0.02% for measurements below 300 K via immersion measurements, and 19.70±0.5% using the measured dimensions and mass. All measured densities were slightly lower than the x-ray-diffraction value, 19.86 g/cm^3 at 294 K.

Thermal-expansion measurements were made in a Netzsch 402 dilatometer with Type-E thermocouple under high purity He-gas atmosphere to minimize surface oxidation and to ensure good thermal contact. We corrected for the thermal expansion of the silica specimen holder. We used a single specimen from the same batch as the RUS specimen used above 300 K, but this sample was not remelted multiple times. Heating and cooling rates of 1 K/min were used. The sample thickness (0.1 cm) was chosen to reduce any through-thickness thermal gradients to realize a uniform transformation throughout the sample volume. A hold time of 1 h was used at the lowest temperatures.

B. RUS measurements

Using resonant-ultrasound spectroscopy, we made measurements using 30–36 resonance peaks to determine the two independent quasi-isotropic effective elastic moduli.[13–15] This method sweeps the frequency driving a broadband transducer in weak contact with the specimen, and detects the macroscopic-vibration resonance frequencies. The elastic moduli are determined by an inverse calculation that uses the specimen mass and dimensions, known crystallographic symmetry, and the resonance frequencies. Errors in specimen geometry, texture, composition (impurities), morphological flaws, voids, and density limit absolute accuracy. From the curvature of the minimum in elastic modulus space during fitting, we determine precisely the 1σ RUS baseline errors (caused by the sum of geometry, morphology, grain size, and texture errors—that is, errors that make the specimen deviate from a homogeneous, isotropic rectangular parallelepiped) to be 0.27% for C_L and 0.17% for G.

For small mass-density errors, RUS measurements can be corrected. Noting that a resonance depends on the moduli and dimensions, as the specimen length and volume change with temperature, or there are slight density errors associated with, for example, voids, the fractional correction to moduli for an isotropic polycrystal material is linearly proportional to the fractional change in length such that $C_{ij}^{\mathrm{corrected}} = C_{ij}^{\mathrm{measured}}(\rho_{\mathrm{x\text{-}ray}}/\rho_{\mathrm{measured}})^{1/3}$. In this way, we can make a simple correction for the difference between our measured density near ambient temperature and the x-ray density. Absolute errors in the RUS-determined moduli arising from density errors at ambient temperature are of order 0.5% if no corrections are made. Correcting to the x-ray density reduces density-induced errors to much less than the RUS baseline errors. The correction we applied was to increase all the high-temperature moduli by 0.51% from those computed using 19.55 g/cm^3 as the density for the RUS fits. Errors introduced by impurities are also much less than the RUS baseline error. Also we made other corrections described in the following.

The RUS measurement cell used a conventional ^{4}He-flow cryostat for temperature control, capable of operation from 2 to 670 K, with the specimen in direct contact with transducers, but otherwise in vacuum above 300 K, and in He exchange gas below. Because ^{239}Pu is a radioactive, self-heating (2 mW/g), toxic fissionable metal, Los Alamos National Laboratory requires compliance with extensive safety precautions. These precautions intrude on our ability to control specimen temperature as precisely as for ordinary materials. However, tests with this RUS system and the nonself-heating ferromagnetic metal Co indicate that our absolute temperature accuracy *at the specimen* is no worse than 2% at 400 K, an error at least an order of magnitude larger than the intrinsic errors for the Type-E thermocouple used in the thermal-expansion measurement, the GaAs diode thermometer used below 300 K for RUS, and the RhFe resistance thermometer used above 300 K for RUS. This error is primarily caused by the necessity of keeping thermometers and specimens separate, and relying on He exchange gas (thermal expansion) and radiation and He gas (RUS) for thermal contact. Noting that near ambient temperature, temperature errors are much less than 1 K, and that the high-temperature specimen exhibited the highest elastic moduli we have ever measured at ambient temperature for pure Pu, as well as minimal observed voids and metallographic imperfections, we corrected the moduli measured below 300 K, based on RUS fits using the geometric density of 19.70 g/cm^3 at 299.3 K by increasing G by 3.5%, and C_L by 2.4% to match the density-corrected high temperature speci-

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Migliori *et al.*: Alpha-plutonium elastic moduli 1995
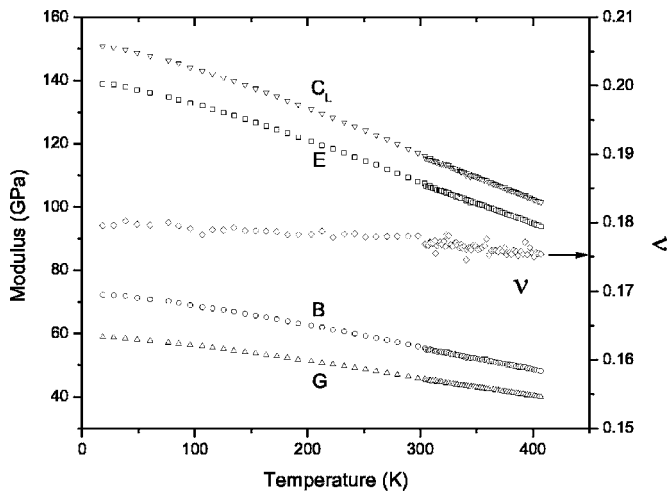
FIG. 1. Temperature dependence of α-plutonium's polycrystal (quasi-isotropic) elastic moduli. The measurement above 300 K was on a different sample than that below 300 K (see the text). Not shown, Eq. (2) fits all the elastic moduli well. Contrary to several previous physical-property studies, we see no evidence for a phase transition. Total 0–400 K change is exceptionally large. The Poisson ratio shows especially unusual behavior; for most metals in this temperature interval, the Poisson ratio *increases* a few percent during warming.

men at 305 K. This is not a large correction for Pu metal, and can easily result from inherent microcracking. That these corrections are different supports the existence of small but detectable morphology problems such as microcracks, common when α-Pu is cooled from the melt.

Unavoidable hysteresis prevents reproducible cooling curves once the first-order phase transition occurs. We find that $T_{\alpha\beta}$, the temperature on warming at which α-Pu transforms to β-Pu determined in the RUS system using an uncalibrated commercial dilatometer sensor, and behavior of the resonances, was 408 K±8 K with a width of about 2 K. The specimen from the same batch, run in the precision dilatometer, transformed at 400 K with a width of about 9 K. The errors are such that these temperatures are easily within the error spread between the two very different measure-
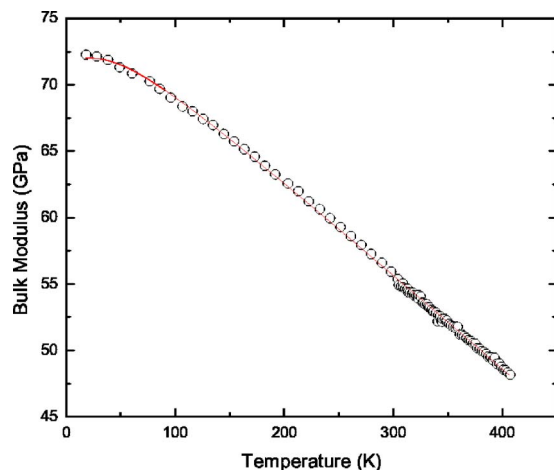
FIG. 2. (Color online) Temperature dependence of α-plutonium's bulk modulus B. Curve represents an Einstein-oscillator model, Eq. (2). The measurement above was 300 K was on a different sample than that below 300 K (see the text). Total change (50%) is unusually large. Low-temperature positive departures, as for G, deserve further study.
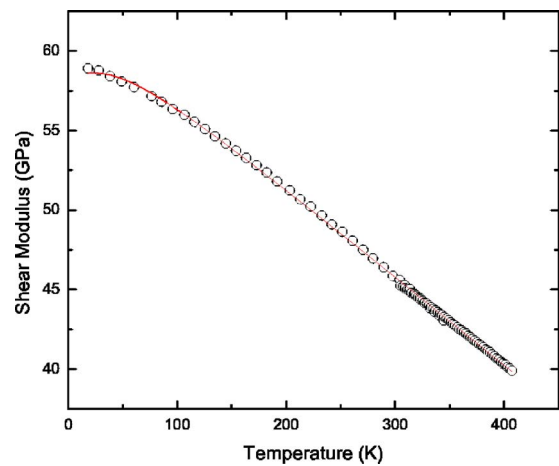
FIG. 3. (Color online) Temperature dependence of α-plutonium's shear modulus G. Curve represents an Einstein-oscillator model, Eq. (2). The measurement above 300 K was on a different sample than that below 300 K (see the text). Total change (47%) is unusually large. Low-temperature positive departures, as for B, deserve further study.

ments systems. Noting that radiation self-heating would make the measured $T_{\alpha\beta}$ appear low, and that in one case He gas was used for heat exchange, while in the other, contact with RUS transducers and radiation were the thermal couplants, we determine $T_{\alpha\beta}$ to be 404 K±4 K in our α-Pu specimen, which was the highest purity and stiffest specimen that we know of, compared to 397.6 K quoted by Wallace,[22] from Ref. 16.

In summary, the errors are less than 2% in temperature at 400 K, less than 0.3% in temperature near 300 K, and, for these particular high-quality specimens, the intrinsic RUS baseline errors are 0.27% for C_L (and about that for B) and 0.17% for G. A 0.3% temperature error (1 K) is equivalent to an error in G of 0.055 GPa, and in B of 0.069 GPa. Thus, at 300 K, the errors are the intrinsic RUS-baseline errors, while at 400 K, the errors are primarily associated with temperature errors and reach about 1.1% for both B and G.

III. RESULTS

Versus temperature, Fig. 1 shows, with all corrections applied, the principal elastic stiffnesses: longitudinal modulus C_L, bulk modulus B, shear modulus G, Young modulus E, and the Poisson ratio:

$$\nu = (1/2)(3B - 2G)/(3B + G). \qquad (1)$$

Figure 2 shows the bulk-modulus results and Fig. 3 the shear-modulus results, together with curves representing Einstein-oscillator-model fits, as described in the following, with fitting parameters in Table I. Table II gives some values of the principal quantities at some temperatures. The graphs

TABLE I. Fitting parameters associated with Eq. (2).

	C_0 (GPa)	s (GPa)	t (K)
C_L (GPa)	150.17	23.19	157.90
B (GPa)	72.01	11.21	156.51
G (GPa)	58.64	8.85	157.06

1996   J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Migliori *et al.*: Alpha-plutonium elastic moduli

TABLE II. Elastic moduli at 18, 298, and 401 K. The density used was from the fitting function used for Fig. 4.

T (K)	C_L (GPa)	B (GPa)	G (GPa)	v	E (GPa)
401	102.27	48.56	40.29	0.1750	94.67
298	117.09	55.93	45.85	0.1781	108.02
18.4	150.82	72.28	58.91	0.1796	138.97

and tables represent results corrected for thermal expansion using measurements shown in Fig. 4, and are consistent with atomic-volume/temperature results reported by Lawson and colleagues.[17] The solid line is a least-squares fit and when this fit is corrected to the x-ray density at 294 K it becomes $\rho(T) = 20.4073 - 1.05575 \times 10^{-3}T - 2.74091 \times 10^{-6}T^2$, where $\rho(T)$ denotes mass density (g/cm^3), and T temperature (K).

IV. DISCUSSION

Table III compares our ambient-temperature measurements with previous ones. Our moduli are higher than all previous measurements by a few percent except for one; probably a result of our unusually pure material, lack of microcracks, and near-theoretical mass density. Only the material used by Laquer is comparable in reported density, but he found an anomalously low bulk modulus!

α-plutonium's shear-modulus/bulk-modulus ratio G/B is remarkably high: 0.82 vs 0.41 for the average metal. Gilman discussed the G/B ratio in terms of ductility, hardness, and the Cauchy discrepancy.[18] Related to high G/B, the polycrystal Poisson ratio is low: 0.18, where, omitting a few aberrant cases, the average value for metals equals 0.31 ± 0.05. As emphasized by Köster and Franz,[19] more than any other
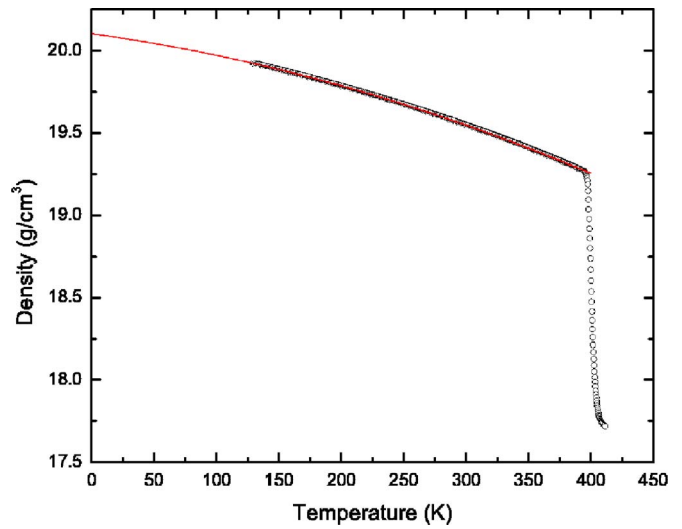
FIG. 4. (Color online) Density of a polycrystal Pu sample from the same batch as was used for the high temperature resonance measurements. The solid line is a least-squares fit. When this fit is corrected to the x-ray density at 294 K it becomes $\rho(T) = 20.4073 - 1.05575 \times 10^{-3}T - 2.74091 \times 10^{-6}T^2$, where $\rho(T)$ is density (g/cm^3) and T is temperature (K).

elastic modulus, the Poisson ratio reflects the character of interatomic-bonding forces. At least for cubic symmetry, these authors (their Fig. 1) showed that when the Poisson ratio falls below about 0.25 the C_{44}/C_{12} ratio (which reflects noncentral interatomic forces) increases strongly. For materials such as beryllium and diamond, a low Poisson ratio results from high resistance of interatomic bonds to bending. But these are high-symmetry materials; in contrast α-plutonium represents a low-symmetry, large-unit-cell ma-

TABLE III. Measured ambient-temperature elastic moduli of polycrystalline α-plutonium.

Source	Year	Method	B (GPa)	E (GPa)	G (GPa)	v
Konobeevsky et al. (Ref. 40)	1958	Resonance	⋯	⋯	⋯	0.21
Bridgman (Ref. 41)	1959	Pressure	50	⋯	⋯	⋯
De Cadenet (Ref. 42)	1960	Pulse	54.3	98.3	40.9	0.199
Gardner and Mann (Ref. 43)	1960	Tensile	⋯	98.7	⋯	⋯
Kay and Linford (Ref. 34)	1960	Resonance	53.4	100.7	42.3	0.186
Lacquer (Ref. 44)	1961	Pulse	46.7	99.3	43.4	0.15
Robinson (Ref. 45)	1961	Pressure	52.2	⋯	⋯	⋯
Andrew and Morgan (Ref. 46)	1966	Pressure	45.7	⋯	⋯	⋯
Cornet and Bouchet (Ref. 47)	1968	Resonance	⋯	99	⋯	⋯
Rosen et al. (Ref. 35)	1968	Pulse	(59.9)[a]	107	44.5	(0.202)
Yemelyanov and Yevstyukhin (Ref. 48)	1969	Handbook	(93)	98	37	(0.32)
Merz et al. (Ref. 49)	1976	Pulse	54.1	(107.5)	46.0	(0.169)
Calder et al. (Ref. 50)	1981	Pulse	(49.4)	92.8	39.1	0.187
Roof (Ref. 51)	1981	Pressure	42	⋯	⋯	⋯
Dabos-Seignon et al. (Ref. 23)	1993	Pressure	43	⋯	⋯	⋯
Present	2006	Resonance	55.9	108.0	45.9	0.178
Average[b]			50.6	100.5	42.9	0.185

[a]Numbers in parentheses indicate computed from other given elastic moduli.
[b]Excluding Ref. 48.

terial. In monoclinic plutonium, a low Poisson ratio can be understood as resulting from several nonequivalent atom positions within the unit cell (eight site symmetries versus one in f.c.c. delta-plutonium). When stress is applied, atoms can change their relative coordinates without changing unit-cell symmetry. In close-packed higher-symmetry structures such as b.c.c. and f.c.c., atoms cannot change their relative coordinates without breaking unit-cell symmetry.

To the measurements in Figs. 2 and 3 we fit an expression based on the assumption that elastic stiffness decreases linearly with the thermal-oscillator-energy increase:[20]

$$C(T) = C_0 - s/(e^{t/T} - 1). \tag{2}$$

Here, C_0 denotes the zero-temperature elastic stiffness and t relates roughly to the average Einstein temperature. Originally an adjustable parameter, in a quasiharmonic Einstein-oscillator model, s becomes[21]

$$s = 2(C_h - C_0) = 3k_B t \gamma(\gamma+1)/V_a. \tag{3}$$

Here C_h denotes the harmonic zero-temperature elastic modulus obtained by linear extrapolation from high temperatures, k_B the Boltzmann constant, γ the Grüneisen parameter, and V_a the atomic volume. From Eq. (2), the high-temperature slope follows as

$$dC/dT = -3k_B \gamma(\gamma+1)/V_a. \tag{4}$$

Solving this relationship for γ and taking the positive root gives $\gamma=5.3$. Solving Eq. (3) gives $\gamma=4.9$. These values differ strongly from the value obtained from Grüneisen's rule: $\gamma=B\beta V_a/C_p=3.0$ (values for V_a and C_p - heat capacity are from Ref. 22; $\beta=1.40\times10^{-4}$ K^{-1}, from our thermal-expansion measurements). We can also compare these values with the Grüneisen parameter obtained from the measured dB/dP,[23] using the approximate relationship $dB/dP=2\gamma+1$, which yields $\gamma=7.0\pm1$. The handbook lattice-specific-heat value is $\gamma=6.8$.[55] Fitting an Einstein-oscillator model to low-temperature thermal-expansion measurements[17] gives $\gamma=3.2$. Temperature dependence of the Debye-Waller temperature gives $\gamma=4.1$.[17] These varying Grüneisen gammas obtained from different physical properties remind one of the more-extreme δ-Pu case where values range from -0.2 to 7 when obtained from various physical properties. There exists an important difference: All the α-Pu values exceed 3.1, thus it does not possess the odd negative Grüneisen parameter and negative thermal expansivity that occur in δ-Pu. α-plutonium's Grüneisen parameter needs further focus. As described in the following, f-electron localization during warming changes the material state and could give a fictitious high Grüneisen parameter.

The fitting parameters in Table I present a big surprise: The t parameter, which relates roughly to the Einstein temperature, is nearly the same for all four considered elastic moduli (B,E,G,C_L): 157 ± 2 K. And large: about 0.76 of the Debye temperature. Varshni's study of 23 various materials yielded much lower t/Θ_D ratios.[20] For example, for ten cubic elements considered by Varshni, considering the elastic shear modulus C_{44}, one obtains $t/\Theta_D=0.43\pm0.15$, about one-half the α-Pu value. In terms of Einstein-Debye phonon models, this implies a phonon density-of-states curve populated un-

usually heavily at higher lattice-vibration frequencies. An alternative interpretation arises: During cooling, electronic changes occur that change the elastic moduli.

The slopes dC/dT figure prominently in melting theories. Usually, focus falls on dG/dT,[24,25] but sometimes also on dB/dT.[26] We can calculate the temperature at which elastic stiffness vanishes (and α-phase melting must occur) by using Eq. (2) and setting $C(T)$ to zero. We obtain

$$T_{C=0} = t/\ln(1 + s/C_0). \tag{5}$$

Solving Eq. (5) gives 1083 K using the bulk modulus and 1117 K using the shear modulus. It seems odd that two independent elastic moduli, one representing resistance to volume change and one representing resistance to shape change, should behave similarly. Also, because B and G represent bounds, the other elastic-stiffness moduli such as the longitudinal modulus, Young (extension) modulus, and Lame' modulus must extrapolate to zero between these temperatures, about 200 K above plutonium's melting point, 915 K.

Zero-temperature elastic moduli provide the best estimate of the Debye temperature Θ_D. Since Einstein's lattice-vibration studies,[27] many authors calculated the Debye temperature from the bulk modulus.[28,29]

$$\Theta_D = K(r_0 B/m)^{1/2}. \tag{6}$$

Here, K denotes a collection of physical moduli, r_0 Wigner-Seitz radius, m atomic mass. Ledbetter showed that the bulk modulus estimates Θ_D very roughly, that one obtains a much better estimate using the shear modulus G.[30] One obtains an "exact" elastic Θ_D from a relationship given by Kim and Ledbetter;[31]

$$\Theta_D = 2933.22 v_m/V_a^{1/3}. \tag{7}$$

Here, v_m denotes mean sound velocity (computed by the familiar Debye-model reciprocal-cube average) and V_a atomic volume. In Eq. (7), units on v_m are cm/μs, on V_a Å^3. The premultiplier contains various physical moduli. This approach gives $\Theta_D=207\pm2$ K. The uncertainty in Θ_D arises mainly from the uncertainty in the shear modulus, where, approximately, $\Theta_D\sim G^{1/2}$. As shown in Table IV, our result is higher than most previous estimates: 118 K from the Lindemann melting equation, 153–200 K from low-temperature specific heat, 178–206 K from elastic moduli. A recent careful specific-heat measurement gave 153 K.[32] We attribute the large specific-heat/elastic modulus disagreement to the difficulty of extrapolating accurately the specific-heat measurements to zero temperature. A recent EXAFS study on α-prime-plutonium (prime means simply the α crystal structure with stabilizing atoms substituted for some of the Pu atoms) yielded a Pu-Pu-bond Debye temperature of 159 K.[33] One expects the corresponding α-plutonium value to be higher. From Table IV, we see that we obtain 203.6 ± 3.6 K by averaging the five highest values: the Sandenaw specific-heat result, the Debye-Waller result, and the Taylor *et al.*, Rosen *et al.*, and present elastic results. Lack of correlation with melting point (Lindemann) and with average vibrational frequency (Einstein) indicates once more plutonium's odd, complex, inconsistent physical-property behavior.

TABLE IV. α-plutonium's Debye temperature by various methods.

Source	Year	Method	Θ_D (K)
Dempsey and Kay (Ref. [52])	1958	Einstein reln.	123
Lacquer (Ref. [44])	1961	Elastic	178
Sandenaw et al. (Ref. [53])	1961	Specific heat	174
Sandenaw (Ref. [54])	1962	Specific heat	200
Gschneidner (Ref. [55])	1964	Lindemann eq.	118
Kay and Loasby (Ref. [56])	1964	Specific heat	192
Lee et al. (Ref. [57])	1965	Specific heat	158
Taylor et al. (Ref. [58])	1965	Elastic	204
Lee et al. (Ref. [59])	1965, 1968, 1970	Specific heat	162
Rosen et al. (Ref. [35])	1968	Elastic	206
Gordon et al. (Ref. [60])	1976	Specific heat	183
Stewart and Elliott (Ref. [61])	1981	Specific heat	177
Lawson et al. (Ref. [17])	1994	Debye-Waller	201
Nelson et al. (Ref. [33])	2003	EXAFS	>159
Lashley et al. (Ref. [32])	2003	Specific heat	153
Present	2007	Elastic	207

The bulk-modulus decrease from 0 to 300 K (and beyond) is particularly large, as shown in Table V, which shows handbook $B(0)/B(300)$ values for several typical metals. Large changes reflect a large Grüneisen parameter. In the α-Pu case, they may reflect also temperature-induced electronic changes among the 5f electrons. Converting itinerant electrons to localized electrons would decrease cohesion and decrease the bulk modulus. The anomalous softening of uranium's bulk modulus arises from the sharp decrease of C_{11} during cooling, caused presumably by magnetic-state changes. The first indication of a large Pu elastic-modulus change with temperature appeared in the coarse ultrasonic elastic-modulus measurements by Kay and Linford.[34] Better resonance-method elastic-moduli measurements by Rosen and colleagues quantified the changes for the Young and shear moduli.[35] Debye-Waller-factor measurements by Lawson and colleagues confirmed this large elastic-modulus change.[17] The Debye-Waller Debye temperature Θ_{DW} connects with the bulk modulus B by the following relationship:

$$B = k_B^2 (\Omega/6\pi^2)^{2/3} \Theta_{DW}^2/\hbar^2. \tag{8}$$

Here, Ω denotes atomic volume, $\hbar$ Planck's constant divided by 2π.

Together with large elastic-moduli changes in the 0–300 K interval, there is another remarkable feature: the very weak decrease of the Poisson ratio v on warming. As described by Slater,[36] the Poisson ratio usually increases with increasing temperature because the material approaches the

TABLE V. Bulk-modulus changes between 0 and 300 K for various materials.

Material	$B(0)/B(300)$
Al	1.075
Cu	1.036
Fe	1.030
Pb	1.104
α-U	0.963
α-Pu	1.288

liquid state where $v = 0.5$, the upper bound for an isotropic solid. For example, in the 0–300 K interval, copper's Poisson ratio increases 1.9%. α-Pu shows a 1.3% *decrease*. As mentioned earlier, the Poisson ratio reflects interatomic bonding. Thus, abnormal Poisson-ratio behavior reflects abnormal interatomic-bonding changes. Especially in δ-Pu (f.c.c.) research, various studies emphasized temperature-induced changes in the localized–delocalized 5f-electron distribution.[37,38] Like δ-Pu, α-Pu shows large elastic softening with increasing temperature.

Elsewhere,[12] we compared our measurement results with theoretical predictions (nearly all zero temperature). Among 14 theoretical studies, none predicted the bulk modulus within 20%. On average, the theories predicted 93±48 GPa, our measured value being 72 GPa.

In Sec. I, we alluded to plutonium's possible covalency and its possible connections to the elastic moduli. Using our present results, we cannot argue convincingly for covalency; however we note two results that suggest covalency: the high G/B ratio and the low Poisson ratio. For seven f.c.c. metals (Ag, Al, Au, Cu, Ni, Pb, Pt) these numbers average 0.30±0.10 and 0.37±0.04. For the two protypical covalent elements (Ge, Si), these numbers average 0.70±0.03 and 0.22±0.01. Our α-plutonium results, 0.82 and 0.18, compare favorably with the covalent values. Related to covalency, in Sec. I we also alluded to possible compound-like behavior. For an element, how can this happen? For a simple crystal structure (f.c.c., b.c.c., c.p.h.), it cannot because all lattice sites share the same Wyckoff site symmetry. However, α-Pu's 16-atom unit cell with 8 different site symmetries means that the plutonium atoms differ locally and thus may bond differently with nearer neighbors. For α-Pu, Zachariasen and Ellinger pointed out that coordinations range from 4 to 13 and 1nn-2nn bond distances range from 2.50 to 3.64 Å.[2] Thus, through its low-symmetry crystal structure, α-Pu shows compound-like behavior. In interpreting α-U's orthorhombic crystal structure, Tucker relied heavily on covalent-bonding mechanisms proposed by Pauling.[39] It is attractive to attribute the covalent and compound-like behavior of these elements to the localized–delocalized 5f-electrons, whose role in actinide bonding remains unsettled.

V. CONCLUSIONS

For α-plutonium, our study produced the following principal results and conclusions:

(1) α-Pu shows a high shear-modulus/bulk-modulus ratio—thus a low Poisson ratio. This may reflect a large-unit-cell, low-symmetry crystal structure that appears in many actinides. It may reflect covalent bonding.

(2) α-Pu shows unusually large elastic-modulus changes in the 0–404 K interval. These may reflect changes in the 5f-electron character (itinerant–localized).

(3) From 18 to 300 K, Poisson's ratio *decreases* by 1.3%, and from 18 to 404 K, it decreases by 2.3%. Such odd behavior indicates interatomic-bonding changes.

(4) From the elastic moduli, we computed a Debye temperature of 207 K. This exceeds most previous estimates, but

agrees within ±4 K with four previous estimates: two from elastic moduli, one from Debye-Waller factor, and one from specific heat.

(5) From dB/dT, we computed a Grüneisen parameter 5.3, unusually high. High Grüneisen parameters reflect high anharmonicity. Differences in Grüneisen parameters obtained from different physical properties probably reflect the failure of simple theory to describe α-Pu's electronic complexity. Also, the dC/dT-derived Grüneisen parameters may be artificially high because of temperature-induced $5f$-electron changes.

(6) Except at the lowest studied temperatures, the $C(T)$ agree well with a simple Einstein-oscillator model. A low-temperature "bump" requires further study.

(7) Einstein-oscillator-model fits to the $C(T)$ yield high Einstein temperatures. These suggest an unusual phonon spectrum.

(8) All the elastic-stiffness moduli extrapolate to zero at the same temperature, near 1100 K, 200 K above the epsilon (b.c.c.)-Pu melting point. From most melting models, one expects the shear modulus to vanish before the bulk modulus.

(9) All the elastic-stiffness moduli show smooth $C(T)$ behavior. This smoothness indicates absence of phase transitions: electronic, magnetic, structural. Particularly, we found no evidence of the 66 K cooperative electron transition proposed by Rosen and colleagues.[35]

(10) We conjectured about possible covalent bonding in actinides and described some covalency/elastic-modulus relationships.

ACKNOWLEDGMENTS

[1]"Gmelin's inorganic crystal structure database," http://barns.ill.fr/dif/icsd/index.html, 2002.

[2]W. H. Zachariasen and F. H. Ellinger, "The crystal structure of alpha plutonium metal," Acta Crystallogr. **16**, 777–783 (1963).

[3]W. Harrison, *Electronic Structure and the Properties of Solids* (Freeman, San Francisco, 1980), Chap. 8.

[4]O. Anderson, *Equations of State of Solids for Geophysics and Ceramics Science* (Oxford University Press, New York, 1995).

[5]D. Wallace, *Statistical Physics of Crystals and Liquids: A Guide to Highly Accurate Equations of State* (World Scientific, River Edge, NJ, 2002).

[6]*Challenges in Plutonium Science*, Los Alamos Science No. 26 (LANL, Los Alamos, 2000), Vols. I, II.

[7]W. B. Holzapfel, "Systematics and anomalies in the equations of states for the lanthanide and actinide elements," Physica B **190**, 21–30 (1993).

[8]F. H. Herbstein, "Methods measuring Debye temperatures and comparison of results for some cubic crystals," Adv. Phys. **10**, 313–355 (1961).

[9]H. Ledbetter, in *Materials at Low Temperatures* (ASM, Metals Park, 1983), Chap. 1, especially p. 37. Here, connections are made between elastic moduli and various physical properties. The Debye temperature follows directly and simply from the elastic moduli and the atomic volume.

[10]B. Lüthi and W. Rehwald, in *Structural Phase Transitions I* (Springer, Berlin, 1981), p. 131.

[11]J. Lashley, A. Lawson, R. McQueeney, and G. Lander, "On the possibility of magnetic moments in plutonium," LAUR Rep. No. 04-3439 (2004).

[12]H. Ledbetter, A. Migliori, J. Betts, S. Harrington, and S. El-Khatib, "Zero-temperature bulk modulus of alpha-plutonium," Phys. Rev. B **71**, 172101–1–4 (2005).

[13]O. L. Anderson, "Rectangular parallelepiped resonance—A technique of resonance ultrasound and its applications to the determination of elasticity at high temperatures," J. Acoust. Soc. Am. **91**, 2245–2253 (1992).

[14]J. Maynard, "Resonant ultrasound spectroscopy," Phys. Today **49**, 26–31 (1996).

[15]A. Migliori and J. Sarrao, *Resonant Ultrasound Spectroscopy* (Wiley-Interscience, New York, 1997).

[16]F. L. Oetting and R. O. Adams, "The chemical thermodynamics of nuclear materials. VIII. The high-temperature heat capacity of unalloyed plutonium metal," J. Chem. Thermodyn. **15**, 537–554 (1983).

[17]A. C. Lawson, J. A. Goldstone, B. Cort, R. I. Sheldon, and E. M. Foltyn, "Atomic thermal vibrations of the light actinide elements," J. Alloys Compd. **213/214**, 426–428 (1994).

[18]J. Gilman, *Electronic Basis of the Strength of Materials* (Cambridge University Press, Cambridge, 2003), Chap. 13.

[19]W. Köster and H. Franz, "Poisson's ratio for metals and alloys," Metall. Rev. **6**, 1–55 (1961).

[20]Y. Varshni, "Temperature dependence of the elastic constants," Phys. Rev. B **2**, 3952–3958 (1970).

[21]H. Ledbetter, "Relationship between bulk-modulus temperature dependence and thermal expansivity," Phys. Status Solidi B **181**, 81–85 (1994).

[22]D. C. Wallace, "Electronic and phonon properties of six crystalline phases of Pu metal," Phys. Rev. B **58**, 15433–15439 (1998).

[23]S. Dabos-Seignon, J. P. Dancusse, E. Gering, S. Heathman, and U. Benedict, "Pressure-induced phase transition in α-Pu," J. Alloys Compd. **190**, 237–242 (1993).

[24]M. Born, "Thermodynamics of crystals and melting," J. Chem. Phys. **7**, 591–603 (1939).

[25]A. N. May, "Extrapolation of the shear elastic moduli of face centred cubic solids to the molten state," Nature (London) **228**, 990–1001 (1970).

[26]J. L. Tallon, "The volume dependence of elastic moduli and the Born-Durand melting hypothesis," Philos. Mag. A **39**, 151–161 (1979).

[27]A. Einstein, "The Planck theory of radiation and the theory of specific heat," Ann. Phys. Chem. **22**, 180–190 (1907).

[28]V. L. Moruzzi, J. F. Janak, and K. Schwarz, "Calculated thermal properties of metals," Phys. Rev. B **37**, 790–799 (1988), Eq. (7).

[29]P. Söderlin, L. Nordstrom, L. Yongming, and B. Johansson, "Relativistic effects on the thermal expansion of the actinide elements," Phys. Rev. B **42**, 4544–4552 (1990).

[30]H. Ledbetter, "Atomic frequency and elastic constants," Z. Metallkd. **82**, 820–822 (1991).

[31]S. Kim and H. Ledbetter, "Low-temperature elastic coefficients of polycrystalline indium," Mater. Sci. Eng., A **252**, 139–143 (1998).

[32]J. Lashley, A. Migliori, J. Singleton, R. McQueeney, R. A. Pereyra, J. L. Smith, and M. S. Blau, "The electron heat capacity of high-purity alpha and stabilized-delta plutonium," JOM **55**, 34 (2003); "Experimental electronic heat capacities of α- and δ-plutonium: Heavy-fermion physics in an element," Phys. Rev. Lett. **91**, 205901–1–4 (2003).

[33]E. Nelson, K. Blobaum, M. Wall *et al.*, "Local structure and vibrational properties of α-Pu martensite in Ga-stabilized δ-Pu," Phys. Rev. B **67**, 224206–1–10 (2003).

[34]A. Kay and P. Linford, "The chemical thermodynamics of nuclear materials VIII. The high-temperature heat capacity of unalloyed plutonium metal," J. Chem. Thermodyn. **15**, 537–554 (1983).

[35]M. Rosen, G. Erez, and S. Shtrikman, "Evidence for a cooperative electron transition in plutonium at low temperatures," Phys. Rev. Lett. **21**, 430–431 (1968).

[36]J. Slater, *Introduction to Chemical Physics* (McGraw-Hill, New York, 1939), p. 240.

[37]A. Freeman and D. Koelling, *The Actinides: Electronic Structure and Related Properties* (Academic, New York, 1974), p. 51.

[38]X. Dai, S. Savrasov, G. Kotliar, A. Migliori, H. Ledbetter, and E. Abrahams, "Calculated phonon spectra of plutonium at high temperatures," Science **300**, 953–955 (2003).

[39]C. W. Tucker, "The crystal structures of metallic uranium," Trans. ASME **42**, 762–770 (1950).

[40]S. Konobeevsky, A. S. Zaimovskii, B. M. Levitskii, Yu. N. Sokurskii, N. T. Chebotarev, Yu. V. Bobkov, P. P. Egorov, G. N. Nikolaev, and A. A.

2000   J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Migliori *et al.*: Alpha-plutonium elastic moduli

Ivanov, in *Some Physical Properties of Uranium, Plutonium, and their Alloys* (United Nations, Geneva, 1958), Vol. **6**, p. 194.

[41]P. Bridgman, "Compression and the α-β phase transition of plutonium," J. Appl. Phys. **30**, 214–217 (1959).

[42]J. DeCadenet, *Plutonium 1960* (Cleaver-Hume, London, 1960), p. 107.

[43]H. Gardner and I. Mann, *Plutonium 1960* (Cleaver-Hume, London, 1960), p. 513.

[44]H. Lacquer, *The Metal Plutonium* (ASM, Metals Park, 1961), Chap. 16.

[45]A. Robinson, reported by H. Gardner, in *Plutonium Handbook* (American Nuclear Society, La Grange Park, 1980), Vol. **I**, p. 59.

[46]J. Andrew and J. Morgan, "Compressibility of oriented alpha plutonium," J. Nucl. Mater. **19**, 115–116 (1966).

[47]J. A. Cornet and J. M. Bouchet, "Elastic properties of Pu in epsilon phase," J. Nucl. Mater. **28**, 303–310 (1968).

[48]*Metallurgy and Metallography of Pure Metals*, edited by V. Yemelyanov and A. Yevstyukhin (Atomizdat, Moscow, 1969).

[49]M. Merz, J. Hammer, and H. Kjarmo, in *Proceedings of the Fifth International Conference on Plutonium and other Actinides*, Baden-Baden, Germany, 1975.

[50]C. Calder, E. Draney, and W. Wilcox, "Noncontact measurement of the elastic constants of plutonium at elevated temperatures," J. Nucl. Mater. **97**, 126–136 (1981).

[51]R. Roof, "Compression and Compressibility studies of Plutonium and a Plutonium-Gallium Alloy," Adv. X-Ray Anal. **24**, 221–230 (1981).

[52]E. Dempsey and A. Kay, "Some investigations on plutonium metal," J. Aust. Inst. Met. **86**, 379–384 (1958).

[53]T. Sandenaw, C. Olsen, and R. Gibney, *Some Physical Properties of Uranium, Plutonium, and their Alloys* (United Nations, Geneva, 1958), Vol. **6**, p. 66.

[54]T. Sandenaw, "Effects of self-induced radiation damage on the low temperature heat capacity of a specimen of electro-refined high purity plutonium," J. Phys. Chem. Solids **23**, 1241–1248 (1962).

[55]K. A. Gschneidner, "Physical properties and interrelationships metallic and semimetallic elements," Solid State Phys. **16**, 275–426 (1964); **16**, 370, 378 (1964).

[56]A. Kay and R. Loasby, "The specific heat of plutonium at high temperatures," Philos. Mag. **9**, 37–49 (1964).

[57]J. A. Lee, K. Mendelssohn, and P. W. Sutcliffe, "Low temperature specific heat of plutonium," Cryogenics **5**, 227 (1965).

[58]J. Taylor, R. G. Loasby, D. J. Dean, and P. F. Linford, *Plutonium 1965* (Institute of Metals, London, 1967), p. 162.

[59]J. Lee, P. W. Sutcliffe, D. J. Martin, and K. Mendelssohn, *Plutonium 1970* (The Metallurgical Society of AIME, New York, 1970), p. 58.

[60]J. E. Gordon, R. O. A. Hall, J. A. Lee, and M. J. Mortimer, "Heat capacities of plutonium and neptunium," Proc. R. Soc. London, Ser. A **351**, 179–196 (1976).

[61]G. Stewart and R. Elliott, *Actinides 1981 Abstract Booklet* (Lawrence Berkeley Laboratory, Berkeley, CA, 1981), p. 206.

Simulation of diagnostic ultrasound image pulse sequences in cavitation bioeffects research

Douglas L. Miller[a] and Chunyan Dou
Department of Radiology, University of Michigan, Ann Arbor, Michigan 48109-0553

Roger C. Wiggins
Department of Internal Medicine (Nephrology), University of Michigan, Ann Arbor, Michigan 48109

Research on cavitational bioeffects of diagnostic ultrasound (DUS) typically involves a diagnostic scanner as the exposure source. However, this can limit the ranges of exposure parameters for experimentation. Anesthetized hairless rats were mounted in a water bath and their right kidneys were exposed to ultrasound. Amplitude modulation with Gaussian envelopes simulated the image pulse sequences (IPSs) produced by diagnostic scanning. A 10 μl/kg/min IV dose of Definity® contrast agent was given during 1–5 min exposures. Glomerular capillary hemorrhage was assessed by histology. A stationary exposure approximated the bioeffects induced by DUS within the beam area. However, the use of five closely spaced exposures more faithfully reproduced the total effect produced within a DUS scan plane. Single pulses delivered at 1 s intervals induced the same effect as the simulated DUS. Use of 100 ms triangle-wave modulations for ramp-up or ramp-down of the IPS gave no effect or a large effect, respectively. Finally, an air-backed transducer simulating DUS without contrast agent showed a zero effect even operating at twice the present DUS guideline upper limit. Relatively simple single-element laboratory exposure systems can simulate diagnostic ultrasound exposure and allow exploration of parameter ranges beyond those available on present clinical systems. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2773991]

I. INTRODUCTION

Diagnostic ultrasound (DUS) imaging utilizes pulsed ultrasound in scanned focused beams, which typically involve temporal average intensities at any fixed point too low for substantial heating and thermal bioeffects. However, bioeffects of nonthermal origin may occur by the mechanism of acoustic cavitation, which depends predominantly on the rarefactional pressure amplitude (RPA) of the pulses. Inertial cavitation involves the growth and inertial collapse of a cavitation nucleus or gas body (stabilized microbubble) in response to RPAs in excess of an inertial cavitation threshold. The threshold is related to the population of cavitation nuclei and their size distribution in the exposed medium. Inertial cavitation by pulsed diagnostic ultrasound appears to be rare, or nonexistent, for intact mammals, because the requisite optimum nuclei are normally rare in tissue. Therefore, diagnostic ultrasound is essentially free of risk for cavitational bioeffects under normal *in vivo* conditions. This general rule is violated by diagnostic ultrasound contrast agents (UCA), which consist of suspended microscopic gas bodies. Injection of these agents into the body radically changes the population of cavitation nuclei from rare to optimum. The commercial introduction of UCA has allowed the study of cavitational bioeffects of diagnostic ultrasound, which was previously possible only for special *in vitro* systems, in mammals. Bioeffects associated with contrast aided diagnostic ultrasound have been the subject of several comprehensive reviews (NCRP, 2002; Dalecki, 2007; Miller *et al.,* 2007a). The microscale bioeffects seen in animal models of human exposure include: perturbation of vascular endothelium; capillary rupture with petechiae; injury of parenchymal cells at the point of capillary rupture, such as cardiomyocyte death in myocardial contrast echocardiography; and glomerular capillary rupture in exposed kidney, leading to renal tubule injury and hematuria.

Early research on *in vivo* bioeffects of ultrasound with contrast agents utilized laboratory exposure systems, which delivered simple pulse trains of single pulses, for example, in testing for hemolysis in mouse heart blood (Dalecki *et al.,* 1997) or petechial hemorrhage in mouse intestine (Miller and Gies, 1998). However, most recent studies of *in vivo* bioeffects potentially associated with contrast-aided diagnostic ultrasound exposure have utilized diagnostic imaging systems for the ultrasound source. The use of a diagnostic scanner allows for imaging the tissue of interest, such as the heart or kidney, and detection of the ultrasound contrast agent (UCA) as it enters the tissue. This assures the exposure of the desired target tissue to diagnostic ultrasound pulses and to the contrast agent. In addition, several specialized UCA modes, which are available on newer machines, may also be tested, such as intermittent scanning. High RPA pulses delivered for an image can destabilize the UCA gas bodies, which then are lost from the tissue image. This gas-body destruction phenomenon is important, because intermittent imaging modes can be used with a varied interval between contrast-destructive images to allow refill of the tissue capillaries with blood containing UCA. With infusion of the UCA, and var-

ied image intervals, the method can reveal tissue capillary perfusion over an organ, such as the heart with suspected ischemia or infarct (Wei *et al.*, 1998). Other methods have been developed specifically for low RPA UCA imaging, but contrast clearance by gas body destruction remains valuable for refreshing an image or imaging at a new position (Averkiou *et al.*, 2003).

For comprehensive cavitation biology research, the use of a diagnostic scanner as the exposure source can limit the range of study parameters. The limitations involve the pulse RPAs, the pulse timing parameters, and the image scan parameters. Many diagnostic scanners have limited RPAs available for exposure, and, in any event, typically do not exceed the guideline upper limit for diagnostic machines. This limit applies to the peak RPA expected *in situ*, divided by the square root of the ultrasound frequency, which is known as the Mechanical Index (MI). The MI is displayed on many ultrasound machines in units of mega pascals (megahertz)$^{0.5}$, and is limited to less than 1.9 (e.g., a 4 MHz imager cannot exceed an RPA of 3.8 MPa anywhere in the scan plane). Timing parameters are limited by the demands of good image formation. Brief pulses are typically used to provide good resolution in the depth dimension, and pulse repetition periods (PRPs) are dictated by the depth of field for echo return. The image pulse sequences delivered to a point are also influenced by the desired image frame rate for minimizing motion artifacts. Trains of single pulses with relevant pulse duration and PRPs can be simulated easily by laboratory equipment. However, DUS image pulse sequences (IPS) are generated by the beam sweeping by a point, for which the RPA ramps-up as the rapidly pulsed ultrasound beam approaches, reaches a maximum, and then declines as the beam passes by the point. Recently, the duration and intermittent spacing of the IPS have been shown to be important, as well as the pulse duration and PRP, for glomerular capillary hemorrhage in rat kidney, even over the limited ranges available on a clinical diagnostic scanner (Miller *et al.* 2007b, c).

The purpose of this present study was to investigate simple methods of utilizing laboratory ultrasonics systems to simulate diagnostic ultrasound exposure, particularly with regard to the IPS. The goal was to achieve greater flexibility in exposure parameters, than is available from actual DUS exposures, for cavitation biology research. Exploration of a non-clinical parameter space is desirable, for example, in order to test IPSs designed for risk mitigation, or to examine the safety of RPAs near or above the guideline maximum to help assure safety and efficacy. An ideal solution of this research problem would be to create a high-power imaging platform with a complete range of selectable and controllable parameters; however such devices are not readily available. Stationary single-element focused transducers, driven by laboratory pulse generators and power amplifiers, allow simulation of diagnostic ultrasound pulse durations and PRPs. However, the targeting of organs, the simulation of scan IPSs, and replicating the exposure coverage of an image scan plane are more difficult. In this study, diagnostic ultrasound guidance for aiming, pulse amplitude modulation for IPS simulation, and multipoint exposure for approximating
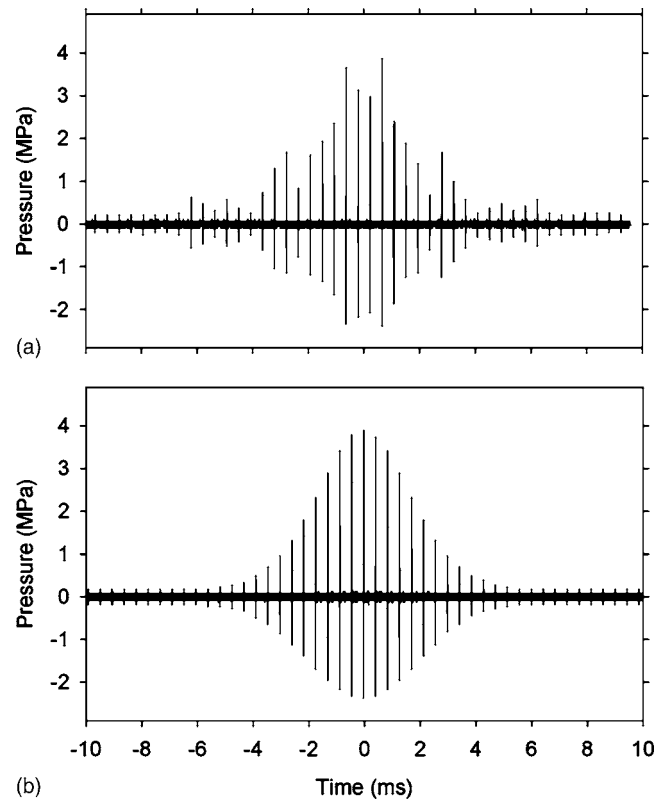
FIG. 1. Oscilloscope traces showing the image pulse sequences for 1.5 MHz diagnostic ultrasound (a) and for a 1.5 MHz damped transducer (b). The pulse peak amplitudes, both compressional and rarefactional, are captured, but the pulse shapes are not resolved at this time scale. The lower trace produced by amplitude modulation of a fixed beam by a Gaussian envelope adequately simulates the diagnostic image pulse sequence produced as the beam scans by the hydrophone element.

scanned areas have been used to partially solve these problems. In a rat model of contrast ultrasound induced glomerular capillary hemorrhage, intermittent 1.5 MHz diagnostic ultrasound scanning of kidney was simulated using a single element transducer. This work demonstrated that it is possible not only to simulate important aspects of diagnostic exposure using relatively simple laboratory systems, but also to examine cavitational bioeffects over a broadened parameter space relative to clinical diagnostic ultrasound.

II. MATERIAL AND METHODS

A. Ultrasound exposure systems

Two exposure systems were used: a diagnostic ultrasound machine and a laboratory exposure system. The diagnostic machine was a GE Vingmed System V (General Electric Co., Cincinnati, OH) with a cardiac phased array probe (FPA2.5). As described previously (Miller *et al.*, 2007c), the probe was clamped in a water bath with the right kidney of a rat located 3.5–4.5 cm from the transducer face. The selected frequency was 1.5 MHz with a 24.4 Hz continuous frame rate, which was modified for exposure by the intermittent frame-trigger feature at a 1 s interval (i.e., a 24.4 Hz IPS generated once per s). The IPS was measured at the focal maximum, as shown in Fig. 1(a). The peak RPA of the pulses was measured as described previously (Miller *et al.* 2007c)

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Miller *et al.*: Image pulse sequences in bioeffects research   2003
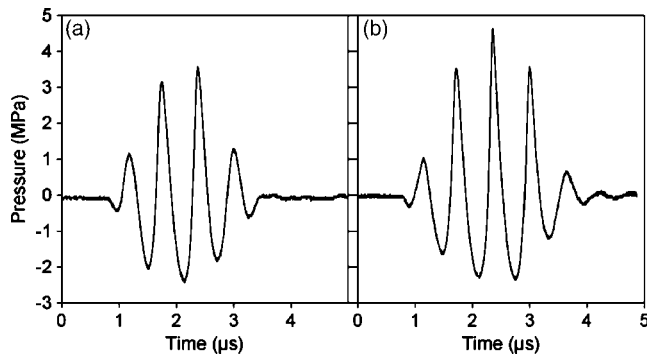
FIG. 2. Oscilloscope traces of the pulse wave forms (a) for the 1.5 MHz diagnostic ultrasound and (b) for the 1.5 MHz damped transducer with the laboratory exposure system. The two pulses are approximately equal in terms of peak rarefactional pressure amplitude and pulse duration.

FIG. 3. Oscilloscope trace showing the 1.55 MHz pulse for the 1.5 MHz air-backed transducer with the laboratory exposure system. This transducer has a somewhat longer duration due to the "ringing" of the transducer element, but was able to generate substantially higher pressure amplitudes than the other transducers (Fig. 2).

and was 2.4 MPa. The pulse wave form is shown in Fig. 2(a), which had a pulse duration of 1.5 μs calculated using the method of the Output Display Standard (AIUM/NEMA, 1992). This measurement was derated to 2.3 MPa (i.e., by 0.45 dB) to account for approximately 1 cm of tissue between the skin surface and the kidney. To help characterize the IPS, the RPA values were fitted with a four parameter Gaussian envelope (SigmaPlot 9.0, Systat Software Inc., Point Richmond, CA), which was used to determine the IPS duration of 4.6 ms as the full width of the envelope at half the peak RPA. The timing parameters needed to characterize the IPS were the pulse duration, PRP, IPS duration, and IPS repetition period (set to a 1 s intermittent interval for this study), which are listed in Table I together with other pertinent parameters.

The laboratory exposure system consisted of a transducer, power amplifier (A-500, Electronic Navigation Industries, Rochester, NY), function generator for generating a pulse train (model 3314A function generator, Hewlett Packard Co., Palo Alto, CA), and an arbitrary wave form generator (model 33220A, Agilent Technologies, Loveland, CO). The function generator was set to simulate the pulse duration and PRP of the diagnostic ultrasound machine using the n-cycle mode with $n=3$ and 0.43 ms PRP (2.3 kHz PRF). The arbitrary wave form generator was used to modulate the amplitudes of the pulse train generated by the function generator, in order to simulate the diagnostic ultrasound IPS. For this purpose, the Gaussian modulation function was used with a 34.7 ms total duration, which gave a 4.6 ms duration IPS. The modulation envelope was triggered at 1 s intervals

internally, but could also be triggered externally, such as from an electrocardiogram signal, if desired. A damped single element transducer (Panametrics A3464, Olympus NDT Inc. Waltham, MA) with 1.9 cm diameter and 3.8 cm focus was normally used for the laboratory system to simulate the diagnostic pulses. The simulated IPS and pulse wave form are shown in Figs. 1(b) and 2(b), respectively, and various parameters are given in Table I. The pulse wave form was slightly longer than the DUS pulse but was a better match than the $n=2$ cycle pulse (1.1 μs duration). Alternatively, a 2.54-cm-diam air-backed transducer with 2.54 cm radius of curvature (built in house) was used for higher RPA exposures. The pulse wave form for this transducer is shown in Fig. 3, with various parameters listed in Table I.

The laboratory exposure transducers did not form an image, and therefore created a problem in targeting the kidney inside the rat. As an aiming aid, the diagnostic ultrasound machine was also used to image the kidney prior to exposure with a high resolution probe, which was then replaced with the laboratory transducer. The overall laboratory system is illustrated in Fig. 4. The kidney was imaged at 8 MHz with default settings and 2.5 cm focus. The laboratory and FPA10 DUS probe were mounted on a horizontal platform with the axes of the transducers parallel and separated by 38.5 mm for the damped transducer and 7 cm for the air-backed transducer. The platform was attached to a gantry, which was used to horizontally translate the exposure transducer into place. The transducer and probe were moved by the fixed offset with a spacer block for exposure in order to place the trans-

TABLE I. Beam and timing parameters for the three ultrasound exposure sources used in this research. The diagnostic ultrasound (DUS) probe and damped transducer were operated at 1.5 MHz, while the air-backed transducer was operated at 1.55 MHz. The beamwidth (BW) for the DUS probe was the thickness of the scan plane, but the diameter of the cylindrical beams for the other transducers. The image pulse sequence (IPS) duration (Dur) was the full width at half maximum of the RPAs fitted with a Gaussian function to simulate the DUS IPS.

Source	Focus (cm)	−6 dB BW (mm)	Pulse Dur (μs)	PRP (ms)	IPS Dur (ms)	Peak RPA (MPa)
DUS FPA2.5	3.8	4.3	1.5	0.43	4.6	2.3
Damped	3.8	3.75	1.7	0.43	4.6	3.2
Air-backed	2.5	1.65	2.2	0.43	4.6	4.8
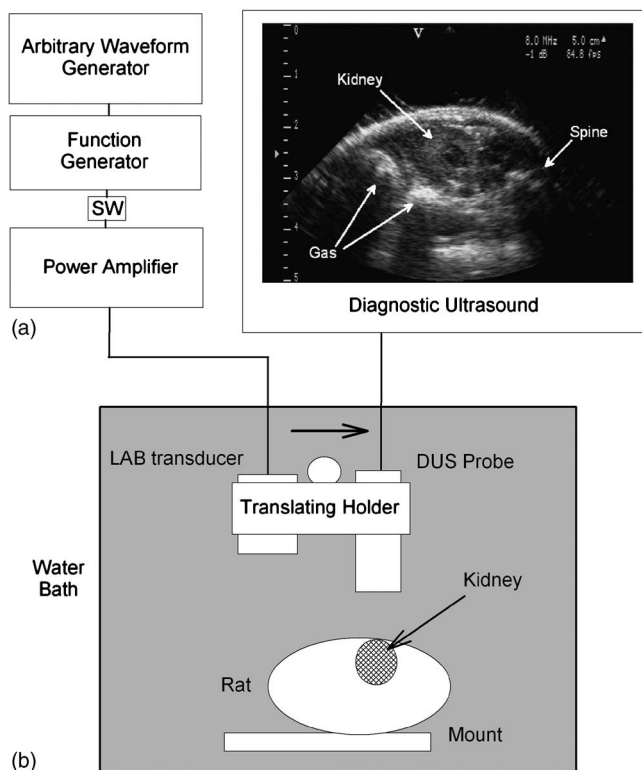
FIG. 4. An illustration of the laboratory exposure system. The view of the water bath is vertically downward, showing the rat positioned on a mounting board on its left side and with its right kidney imaged by 8 MHz diagnostic ultrasound (DUS). The DUS image was used to locate the kidney and aim the scan plane at the middle of the kidney (indicated by the appearance of a darker, low echo region in the center). The laboratory (LAB) transducer, driven by the associated electronic instruments, was translated horizontally into place for exposure of the kidney, which was controlled manually by a switch (SW). This aiming scheme was helpful to avoid the bowel gas in the intestines and the spine, which block the ultrasound transmission (note the acoustical shadows distal to these features).

ducer focus at the point of the DUS probe focus in the image. The purpose of the replacement method was to assure that the exposure transducer had a clear path to the center of kidney, between the spine and intestines, as visualized by the DUS scan (Fig. 4).

The single element damped transducer provided a good simulation of the DUS pulse, PRP, and IPS. However, because the beam was fixed, only one beam line was exposed through the kidney, which was only a fraction of the volume exposed by the DUS cross-sectional scan plane. In two tests performed to simulate the scanned volume, the fixed beam damped transducer was used to make three or five separate exposures within a simulated scan plane. Each exposure duration was the same 1 min duration used for the DUS exposure. For the three-exposure method, the parallel beam positions were separated by 3.75 mm (the −6 dB beamwidth), which produced a total −6 dB width of 11.25 mm covering most of the kidney. For the five-exposure method, the parallel beam positions were separated by 1.88 mm (half the −6 dB beamwidth) to provide a more uniform exposure, which produced a similar total −6 dB width of 11.27 mm. The use of a spacing smaller than half the −6 dB beamwidth was not tried because it would greatly increase the total exposure time, without substantially improving the exposure uniformity.

In addition, the arbitrary wave form generator was used to generate several other amplitude-modulation envelopes. Square wave envelopes of different durations were used to generate bursts of pulses, including 1, 2, and 12 pulses (5 ms envelope, which was roughly the same duration as the DUS Gaussian envelope). These square wave bursts eliminated the ramp-up and ramp-down of the RPAs in the IPS. As noted previously (Miller *et al.*, 2007b), a relatively long duration ramp-up time tends to mitigate the bioeffects, apparently by destroying the gas bodies before the peak RPA pulses arrive. To test the influence of long ramp duration, a 100 ms Gaussian envelope was also tested, which simulated a very slow DUS image frame rate such as can be achieved with Doppler imaging. Ramp-shaped IPS envelopes were also tested, which were 100 ms duration either for ramp-up or for ramp-down using triangle-wave modulation. The ramp-up IPS consisted of a pulse train with linearly increasing RPA for 100 ms (about 232 pulses) to the peak RPA of 2.3 MPs, while the ramp-down was the temporal reverse of this sequence. For this pair of tests, the ramp-up was hypothesized to minimize the bioeffect, whereas the ramp-down was expected to induce maximal bioeffect. The simple ramp-up and ramp-down IPS envelopes were not possible for the DUS machine, except approximately for the trailing and leading edges of a scan window, respectively.

B. Animal preparation and methods

This investigation was conducted with the approval of the University Committee on the Use and Care of Animals, University of Michigan. Hairless rats (CD hairless, Charles River Laboratories, Wilmington, MA) were anesthetized by intramuscular injection of pentobarbital at a dose of 50 mg/kg. A 24 gauge cannula was inserted into a tail vein for UCA injections. The rats were mounted on their left side on a plastic holder, which was held vertically in a bath of 37 °C degassed water. A total of 80 rats were used in the study with 75 yielding usable data (5 rats suffered anesthetic death or other problems, and were excluded from the study). Fifty-four rats were exposed using the damped laboratory transducer in 12 groups, and 15 were exposed using the air-backed laboratory transducer in three groups. In addition, the data from exposure response tests in the previous study with DUS (Miller *et al.*, 2007c) were used in this study for comparison to the newer data. Those data were taken with a tissue mimicking attenuator placed between the probe and the rat and were augmented by data from this study for six rats exposed by the DUS system without the attenuator.

A clinical diagnostic UCA (Definity®, Bristol-Myers Squibb Medical Imaging, Inc., N. Billerica, MA) was prepared each day according to the manufacturer's instructions. The methods have been described in detail previously (Miller *et al.*, 2007c). Sixty μl of the agent was diluted in a syringe containing 3 ml of sterile saline and mounted in a syringe pump. With the rat mounted in a water bath, a 30 cm extension tube was filled with the diluted agent and used to connect the syringe to the tail vein cannula. The agent was then infused at 10 μl/kg/min (0.5 ml/kg/min of diluted agent) with exposure beginning at 15 s after starting the in-

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Miller *et al.*: Image pulse sequences in bioeffects research   2005

fusion (to allow agent to begin circulating). The DUS exposures were 1 min in duration. The one, three, and five position exposures with the laboratory system were 1, 3, or 5 min in total duration (1 min at each point to simulate the DUS dwell time) with total UCA infusions of 10, 30, or 50 μl/kg, respectively, during exposure. Sham exposures consisted of ultrasound exposure without the contrast agent in the circulation.

The rats were removed from the water bath and mounting board after contrast ultrasound exposure, and the kidney was removed after approximately 5 min. The primary end point was the percentage of glomeruli with red blood cells in the Bowman's space counted in histological sections, as described previously (Miller *et al.*, 2007c). Histological processing, sectioning at 5 μm, and hematoxylin and eosin staining were performed at the Research Histology and Immunoperoxidase Laboratory of the University of Michigan Comprehensive Cancer Center Tissue Core. Slides were scored blind using a 20$\times$ objective on a light microscope (model DMRB, Leica Inc., Deerfield, IL). The percentage of glomeruli with capillary hemorrhage was determined in two ways: either for the entire section or for only the beam entrance area within the renal cortex (the layer containing the glomeruli). For the percentage over entire sections, the number of glomeruli with red blood cells in Bowman's space was counted over the entire cross-sectional area and divided by the total number of glomeruli in the section, times 100. For the percentage within the fixed beam entry area of the cortex, the thickness and total area of the cortex were determined using image analysis software (Spot Software v4.5, Diagnostic Instruments Inc., Sterling Heights, MI), and the beam entry area was calculated as the cortex thickness times the -6 dB beam diameter. The total number of glomeruli within the beam entry area was estimated as the total number counted in the entire section multiplied by the ratio of the beam entry area to the total cortex area. The percentage of glomeruli with capillary hemorrhage within the beam area was then calculated as the number counted on the beam entrance side of the kidney, divided by the estimated total number within the beam entry area, times 100. This method was used for consistency in accounting for the area exposed by the fixed beam, but it arbitrarily presupposes that the effected glomeruli were all within the -6 dB beamwidth, which was not true for some conditions and gave some results in excess of 100%.

Numerical results are presented as the mean plus/minus one standard deviation, or plotted as the mean with standard error bars. For statistical analysis, Student's t-tests or Mann-Whitney rank sum tests were used as appropriate to compare means of the measured parameters. Statistical significance was assumed at $P=0.05$.

III. RESULTS

The exposure response of the glomerular hemorrhage effect was compared for full-section scoring of the DUS exposures and entry area scoring of the damped laboratory transducer exposure at one position. Results are shown in Fig. 5. The laboratory transducer was operated at 1.6, 2.3,
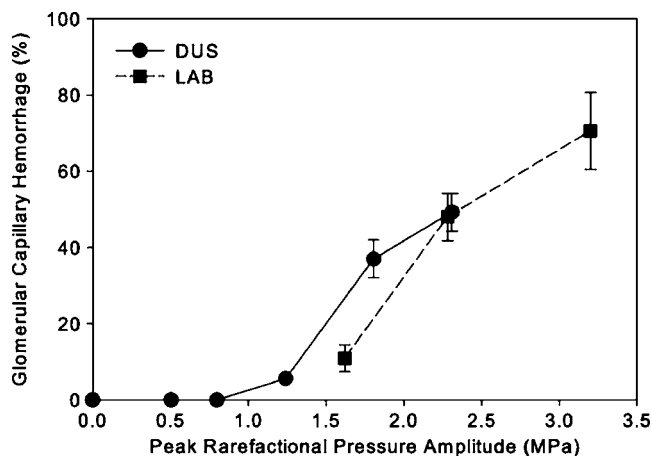
FIG. 5. A comparison of results for the diagnostic ultrasound (DUS, circles) and damped laboratory transducer (LAB, squares) exposures. The glomerular capillary hemorrhage was determined for the entire sections for the DUS, but only for the entry beam area for the LAB system. The guideline upper limit for 1.5 MHz DUS would correspond to about 2.3 MPa on this plot, which precluded use of the DUS system at higher exposure RPAs (e.g., 3.2 MPa from the LAB system).

and 3.2 MPa and extended the exposure response to an equivalent MI value of 2.6 (i.e., greater than the DUS guideline upper limit of 1.9). The two sets of data agree well at 2.3 MPa, but the single element was somewhat less effective for the 1.6 MPa point. The entry beam area only included about 0.16 of the cortex, which contained only about 45 of the total 280 glomeruli on a section. Although the percentage was adjusted for the reduced area, a reduced efficacy might be expected because only the central axis of the beam area was exposed to the peak RPA (as opposed to the entire central plane for the DUS scan).

The results for the multipoint simulation of a scanned DUS are graphed for a peak RPA of 2.3 MPa in Fig. 6. All the exposed samples were evaluated over the entire sections. The single fixed beam exposure produced much less effect than the DUS exposure ($P<0.01$) as expected from the re-
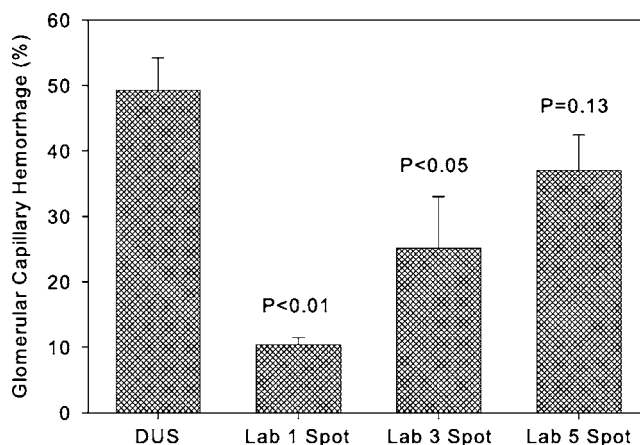
FIG. 6. Results for glomerular capillary hemorrhage induced by the DUS scan, compared to fixed beam exposure at one, three, or five spots, all at a peak RPA of 2.3 MPa. These were scored as the percentage of glomerular hemorrhage over the entire histological sections. Statistical P values are given for the multipoint exposure relative to the DUS exposure. The multiposition exposures provide a better approximation of the scanned exposure bioeffects determined over the entire sections.

Miller *et al.*: Image pulse sequences in bioeffects research
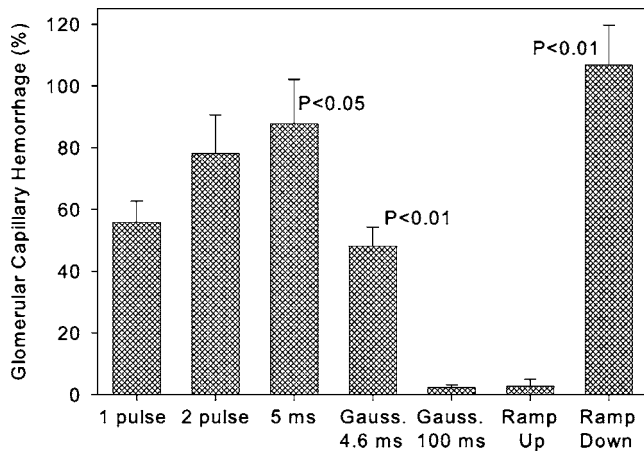
FIG. 7. Results for glomerular capillary hemorrhage in the beam entry area for different pulse sequence envelopes using the damped laboratory transducer at a peak RPA of 2.3 MPa. Statistical P values are for comparisons of the two groups indicated. The square envelopes, which gave bursts of pulses at 1 s intervals, produced an apparently increasing effect with increasing numbers of pulses, but this increase was not statistically significant. A long 100 ms Gaussian envelope, which simulates a slow image frame rate, eliminated the bioeffect seen for a 4.6 ms Gaussian envelope, which confirmed a previous observation of the mitigation of this bioeffect for slow frame rate Doppler DUS. This mitigation strategy for the bioeffect was also confirmed using two similar but reversed ramp envelops: The 100 ms ramp-up envelope gave no significant effect, while a large ($\sim$100%) effect was seen for the ramp-down envelope.

duced area exposed, and the effect elicited by exposure at three positions was also significantly less than the DUS exposure ($P<0.05$). However, the result for the five-position exposure was not significantly different ($P=0.13$), indicating that multipoint fixed beam exposures can approximate the bioeffects induced by scanned DUS exposures.

Glomerular capillary hemorrhage was determined for several fixed beam exposures with different IPS envelopes using a RPA of 2.3 MPa. Results are shown in Fig. 7 for evaluation of the entry beam area. For square modulation, which generated a burst of pulses of equal RPA, there was no significant difference between 1 pulse, 2 pulses, or 12 pulses, only a trend toward greater effect. The 12 pulse-burst (5 ms envelope) exposure produced significantly greater effect ($P<0.05$) than the 4.6 ms Gaussian envelope.

In a test of the variation of this bioeffect obtained by slowly ramping-up the RPA in a pulse sequence, the 100 ms Gaussian envelope was significantly less effective than the 4.6 ms envelope ($P<0.01$) and not significantly different from a zero result in shams ($P=0.15$). Results are shown in Fig. 7 for the beam entry scoring method. Likewise the 100 ms ramp-up envelope did not produce a statistically significant effect. However, the 100 ms ramp-down envelope, which was the temporal reverse of the ramp-up envelope, produced a significant ($P<0.01$) and impressive effect of approximately 100%.

As noted previously, sham exposure with the DUS system did not produce any discernible glomerular capillary hemorrhage (Miller *et al.*, 2007c). This result presumably indicates a lack of cavitation nuclei accessible (i.e., with a sufficiently low cavitation threshold) to the diagnostic ultrasound pulses. To extend this test to higher peak RPAs, an air

backed laboratory transducer was used to expose rats with no injection, with infusion of the saline vehicle only and with the UCA infusion. This exposure simulated the DUS IPS, but at a RPA substantially above the normal diagnostic limits. The saline vehicle, though sterile, was hypothesized to contain some cavitation nuclei, which might nucleate cavitation at the 4.8 MPa peak RPA (equivalent to MI=3.9). However, the result for the control rats and those injected with the saline vehicle were all zero. The rats injected with UCA, as expected, had a significant effect ($P<0.01$), which was $12.2\pm1.7\%$ of the entire cross-sectional area, or $184\pm30\%$ of the beam entry area.

IV. DISCUSSION AND CONCLUSIONS

In cavitation biology research related to medical applications of ultrasound contrast agents, the use of laboratory exposure systems can be of value, particularly to explore ranges of parameters not available on commercial diagnostic ultrasound machines. As noted in Sec. I, such research could, for example, test strategies for mitigation of known bioeffects, or explore conditions above the guideline upper limits of DUS for potential bioeffects. In this study, single element laboratory transducers were used to simulate the pulse sequences produced during DUS imaging (Fig. 1). Gaussian pulse amplitude modulation envelopes were used to yield pulse RPA sequences similar to the DUS IPS, and also to examine pulse burst (square envelope), long duration Gaussian modulation and ramp-up or down envelopes.

The use of a single aiming point for the stationary beams appeared to simulate the bioeffects induced by DUS with contrast agent (Fig. 5), with some reduction in sensitivity under near threshold conditions. The use of several exposures at closely spaced positions to more faithfully reproduce the exposure produced by a DUS scan generated comparable bioeffects even for evaluation over the entire kidney cross section (Fig. 6).

Square RPA modulation was used to produce one or more pulses in bursts. Single pulses with a 1 s PRP produced the same effect as the Gaussian envelope simulation of a 4.6 ms DUS IPS (Fig. 7). This indicates that simple (no modulation envelope) exposure with widely spaced single pulses can approximate the effect of the complex DUS exposure to some extent. Use of more pulses (2 or 12) did not generate a statistically significant increase in the effect. However, the 5 ms (12 pulse) exposure generated significantly higher effect than the 4.6 ms Gaussian simulation.

RPA modulation at different rates was investigated to confirm the observation that a slow Doppler imaging mode mitigated the production of bioeffects seen for B mode imaging with shorter IPS durations (Miller *et al.*, 2007b). A 100 ms Gaussian envelope eliminated the bioeffects seen for the 4.6 ms Gaussian envelope used to simulate the DUS exposure. The hypothesis that this mitigation was due to the relatively long duration ramp-up of pulse RPAs was confirmed by using two similar, but time-reversed RPA ramp (triangle-wave) modulation envelopes. The envelope giving a ramp-up of RPA to the peak produced no significant effect,

while the identical set of pulses delivered with a ramp-down envelope generated the largest effect seen in the set of tests (Fig. 7).

Finally, an air-backed transducer giving a relatively high RPA of 4.8 MPa was used to simulate exposure from a DUS scanner operating well above the guideline upper limit (approximately twice the present USFDA limit). This produced no effect for control exposure without injections, and the expected large effect, which extended well beyond the −6 dB beam diameter, for the UCA infusion. However, this exposure also produced no effect for injection of the saline vehicle, which might have been expected to enhance the cavitation nuclei population for cavitation nucleation at the relatively high peak RPA. This result emphasizes the large departure from normal cavitation nuclei populations, which is caused by the use of UCA.

Single element laboratory exposure systems were used to simulate diagnostic ultrasound exposure for cavitation bioeffects research using UCA nucleation. The glomerular hemorrhage induced by contrast aided diagnostic ultrasound in rats was approximated by use of fixed-transducer exposure aimed with the aid of a diagnostic ultrasound image. The entire effect over the scan plane was approached by using multiple closely spaced exposures. Unusual triangle-wave modulation was used to show mitigation of bioeffects for slow ramp-up of the pulse RPAs. RPAs above the upper limits for DUS were tested with both damped and air-backed laboratory transducers simulating DUS image pulse sequences. Thus, relatively simple single element laboratory exposure systems can simulate ultrasound exposure from diagnostic scanners in cavitation biology research and allow exploration of parameter ranges beyond those available on present clinical systems.

American Institute of Ultrasound in Medicine and the National Electrical Manufacturers Association (AIUM/NEMA). (**1992**). "Standard for real-time display of thermal and mechanical acoustic output indices on diagnostic ultrasound equipment," Rockville, MD, pp. 1–58.

Averkiou, M., Powers, J., Skyba, D., Bruce, M., and Jensen, S. (**2003**). "Ultrasound contrast imaging research," Ultrasound Q. **19**, 27–37.

Dalecki, D. (**2007**). "WFUMB safety symposium on echo-contrast agents: Bioeffects of ultrasound contrast agents *in vivo*," Ultrasound Med. Biol. **33**, 206–213.

Dalecki, D., Raeman, C. H., Child, S. Z., Cox, C., Francis, C. W., Meltzer, R. S., and Carstensen, E. L. (**1997**). "Hemolysis in vivo from exposure to pulsed ultrasound," Ultrasound Med. Biol. **23**, 307–313.

Miller, D. L., and Gies, R. A. (**1998**). "Gas-body-based contrast agent enhances vascular bioeffects of 1.09 MHz ultrasound on mouse intestine," Ultrasound Med. Biol. **24**, 1201–1208.

Miller, D. L., Averkiou, M. A., Brayman, A. A., Everbach, E. C., Holland, C. K., Wible, J. H., and Wu, J. (**2007a**). "Bioeffects considerations for diagnostic ultrasound contrast agents," J. Ultrasound Med., in press.

Miller, D. L., Dou, C., and Wiggins, R. C. (**2007b**). "Doppler mode pulse sequences mitigate glomerular capillary hemorrhage in contrast aided diagnostic ultrasound of rat kidney," IEEE Trans. UFFC., in press.

Miller, D. L., Dou, C., Wiggins, R. C., Wharram, B. L., Goyal, M., and Williams, A. R. (**2007c**). "An *in vivo* rat model simulating imaging of human kidney by diagnostic ultrasound with gas-body contrast agent," Ultrasound Med. Biol. **33**, 129–135.

National Council on Radiation Protection and Measurements (NCRP). (**2002**). "Exposure criteria for medical diagnostic ultrasound: II. Criteria based on all known mechanisms," *Rep. No. 140*, Bethesda, MD, 574 pp.

Wei, K., Jayaweera, A. R., Firoozan, S., Linka, A., Skyba, D. M., and Kaul, S. (**1998**). "Quantification of myocardial blood flow with ultrasound-induced destruction of microbubbles administered as a constant venous infusion," Circulation **97**, 473–483.

Wigner distribution of a transducer beam pattern within a multiple scattering formalism for heterogeneous solids

Goutam Ghoshal and Joseph A. Turner[a)]
Department of Engineering Mechanics, W317.4 Nebraska Hall, University of Nebraska-Lincoln, Lincoln, Nebraska 68588-0526

Richard L. Weaver
Department of Physics, University of Illinois at Urbana-Champaign, 1110 W Green Street, Urbana, Illinois 61801

Diffuse ultrasonic backscatter measurements have been especially useful for extracting microstructural information and for detecting flaws in materials. Accurate interpretation of experimental data requires robust scattering models. Quantitative ultrasonic scattering models include components of transducer beam patterns as well as microstructural scattering information. Here, the Wigner distribution is used in conjunction with the stochastic wave equation to model this scattering problem. The Wigner distribution represents a distribution in space and time of spectral energy density as a function of wave vector and frequency. The scattered response is derived within the context of the Wigner distribution of the beam pattern of a Gaussian transducer. The source and receiver distributions are included in the analysis in a rigorous fashion. The resulting scattered response is then simplified in the single-scattering limit typical of many diffuse backscatter experiments. Such experiments, usually done using a modified pulse-echo technique, utilize the variance of the signals in space as the primary measure of microstructure. The derivation presented forms a rigorous foundation for the multiple scattering process associated with ultrasonic experiments in heterogeneous media. These results are anticipated to be relevant to ultrasonic nondestructive evaluation of polycrystalline and other heterogeneous solids.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2773989]

I. INTRODUCTION

Acoustic propagation in random media at wavelengths comparable to or greater than the size of typical heterogeneities is widely studied, often with a view toward characterization of the microstructure. At low frequencies, such that the wavelengths are very long compared with the length scales of the heterogeneity, the medium is an effectively homogeneous continuum. Measurements of acoustic wave speeds, attenuations, and allowed acoustic polarizations can provide a means for material characterization,[1–7] though microstructural features are not resolved. In such media the chief source of coherent wave attenuation is dissipative and generally unrelated to the microstructures of interest. At moderate frequencies, attenuation is augmented by diffuse scattering[8–10] out of an acoustic beam leading to the possibility of microstructure characterization by means of the frequency dependence of acoustic velocities and attenuation. At slightly higher frequencies where one expects to find significant dependence on microstructure, there is much less literature. This is largely traceable to the very high mean field attenuations characteristic of this frequency range; coherent propagation is too weak for measurement except in special circumstances.[11]

Diffuse multiply scattered fields, however, are not adversely impacted by strong attenuation of the coherent propagation. After many random scatterings or reflections, an acoustic field has lost its original phase coherence, and the field variables become random, with zero mean. The mean square of the field, related to the mean wave energy density, however, remains a meaningful quantity. Diffuse field studies have successfully characterized sources, fields, and media by means of measurements of the evolving wave energy.[12–17] Ultrasonic diffuse field studies have shown themselves capable of characterization of internal friction in polycrystalline microstructure and associated grain noise,[13–17] revealing the Anderson localization of the modes of a disordered structure,[18] and characterization of ultrasonic sources by means of the spectral distribution of the resulting diffuse wave energy.[19,20]

Of particular interest here are the ultrasonic studies of diffuse waves in multiply scattering elastic media. They may be classified in two groups. One concerns work, theoretical[21] and experimental,[17] in which the received signal is modeled as consisting entirely of singly scattered contributions. The other assumes, and/or models, the received signal as thoroughly multiply scattered, i.e., as a fully developed diffuse field.[16,22–26] The former case presumably applies at times sufficiently soon after a transient source has acted, while the latter case applies at times sufficiently long after that source has acted. A clear distinction between the two regimes obvi-

ously requires that studies be carried out in the time domain. But it also demands that we understand the transition between the regimes, a topic excluded from previous investigations. The first of these demands regarding the time domain analysis has not been met. The second, regarding the transition regimes, is an inconvenience at best, and at worst (if signals at early times are contaminated by electronic cross talk or double scattering, and if late times are inaccessible because of strong absorption) will prohibit application of these diffuse energy methods entirely. For this reason a theory for the transition regime is needed.

Several years ago, it was recognized that diffuse ultrasonic energy, at all times from the singly scattered regime to the fully developed regime, ought to be governed by equations like those already known for the transport of diffuse optical energy through multiply scattering atmospheres[27–32] and the transport of neutrons through reactors,[33–35] that is, an equation of radiative transfer. In a series of theoretical papers, Turner and Weaver[22,23,21] in an ultrasonic context, Papanicolaou and co-workers,[36] and van Tiggelen and co-workers[37,38] for seismic waves and elastic waves,[39,40] developed equations of radiative transfer (RTE) for elastic waves in multiply scattering media and argued by consideration of energy flow and conservation.

The form that these equations take is instructive. The relevant dependent variable is not energy, or even energy flux, but *specific intensity*. For a scalar wave this quantity has a single component, $I(\mathbf{x},\hat{\mathbf{n}},t,\omega)$, which may depend on position $\mathbf{x}$, time t, frequency ω, and direction of propagation $\hat{\mathbf{n}}$. I represents the energy flux in a frequency band of unit width, per unit solid angle in direction $\hat{\mathbf{n}}$ at frequency ω at time t and position $\mathbf{x}$. The scalar RTE for I is given by

$$\frac{\partial I(\mathbf{x},\hat{\mathbf{n}},t)}{\partial t} + c\hat{\mathbf{n}} \cdot \nabla I(\mathbf{x},\hat{\mathbf{n}},t) + \kappa I(\mathbf{x},\hat{\mathbf{n}},t)$$

$$= \int p(\hat{\mathbf{n}},\hat{\mathbf{m}})I(\mathbf{x},\hat{\mathbf{m}},t)d^2\hat{\mathbf{m}}, \tag{1}$$

where the dependence on frequency, ω, has been suppressed.

Equation (1) indicates that I is transported at a speed c in direction $\hat{\mathbf{n}}$, attenuates at a rate (per unit time) κ, and is augmented by scattering from direction $\hat{\mathbf{m}}$ into direction $\hat{\mathbf{n}}$. The scattering function p is characteristic of the microstructure. The attenuation, κ, will have contributions from both scattering and dissipation.

For an electromagnetic wave[41,27] or an elastic wave[22] the quantity I has Stokes components, representing the specific intensity in different polarizations. In this case, the scattering function p becomes a Stokes matrix. For elastic waves there are five Stokes components. Equations for elastic wave transport of this form have been derived by Weaver,[16] by Turner and Weaver,[22,23] and by Ryzhik *et al.*[36]

Solution methods, established for applications to optical and neutron transport, are most developed for the planar-independent case corresponding to a steady state planar source of specific intensity. Solutions are obtained with spatial dependence confined to a single dimension. Such limits allow application to a wide class of problems encountered in stellar and planetary atmospheres. In ultrasonics, however,

where measurements must be carried out in the time domain, and where insonification is rarely a plane wave, such limits are inapplicable. Numerical solutions of elastic wave RTE have been obtained by Turner and Weaver[22,23,21] for the steady-state case and for the time domain but have so far been confined to the case of planar insonification. More work is needed.

It is not obvious, however, how an equation like Eq. (1) should be applied to the case of insonification and detection by realistic piezoelectric transducers. It is not difficult to insert, in Eq. (1), a source term corresponding to a pencil beam or a spherical wave in an attempt to model a transducer, but real transducers do not generate intensity in that form. In addition, it is not clear what errors might be induced by so doing. By reciprocity, the same uncertainty applies to the modeling of detection. Furthermore, and perhaps more disturbing, Eq. (1) lends itself with difficulty to a rational interpretation (notwithstanding that it has been derived reasonably). In particular, the dependent variable I is ostensibly defined at arbitrary time and frequency, and at arbitrary position and propagation direction. However, these are complementary independent variables, and any prescription for simultaneous dependence on t and ω, or $\mathbf{x}$ and $\mathbf{n}$ (because propagation direction $\mathbf{n}$ corresponds to the direction of a wave vector $\mathbf{k}$ in spatial Fourier transform space) is suspicious. One might suspect that Eq. (1) involves an approximation akin to those employed in ad hoc signal processing schemes,[42] which attempt simultaneous decomposition in time and frequency. In any case it is clearly appropriate to revisit the derivation of Eq. (1), but with care taken to insert the transducers themselves into the model from the start. This is the purpose of the present communication.

In Sec. II the problem of mean and mean square signals from model source and receiver distributions in a random medium is formally posed. It is solved by familiar perturbative techniques, involving expansions to leading order in fluctuating material properties. The mean signal (or mean square signal) is shown to be equal to a certain convolution between the mean Green's function (or Green's covariance) and the model transducer functions. A Dyson equation is obtained for the mean Green's function, and a Bethe-Salpeter equation is obtained for the Green's function covariance. The latter equation is expanded in a multiple scattering series and the results inserted into the previously derived convolution expression for the mean square signal. It is found that the mean square signal from the receiver is expressible as a multiple scattering series. Each term of that series involves a concatenation of generalized Wigner distribution functions.[43] By comparison with the RTE it is found that the mathematics indicates that the potentially irrational concept "specific intensity" is actually a rational concept, a Wigner transform. There are no ad hoc prescriptions for simultaneous time and frequency distributions; the mathematics clearly indicates that it is the Wigner function which plays the role of specific intensity. It furthermore shows that transducers are not to be represented by the specific intensity they generate, but rather the more precise concept, their field's Wigner distribution.

II. MATHEMATICAL PRELIMINARIES I: DYSON EQUATION AND THE MEAN GREENS' FUNCTION

We begin with the governing partial differential equation (PDE) for the Green's function of an elastic medium with constant material density (set to unity) and modulus that varies randomly in space

$$\left[- \delta_{li} \frac{\partial^2}{\partial t^2} + \frac{\partial}{\partial x_k} \mu_{klij}(\mathbf{x}) \frac{\partial}{\partial x_j} \right] G_{i\alpha}(\mathbf{x}, \mathbf{y}, t) = \delta^3(\mathbf{x} - \mathbf{y}) \delta(t) \delta_{l\alpha}.$$
(2)

The position-dependent modulus tensor $\mu_{klij}(\mathbf{x})$ has a mean value of C^0_{klij}, and fluctuations away from the mean given by $\gamma_{ijkl}(\mathbf{x}) = \mu_{ijkl}(\mathbf{x}) - C^0_{ijkl}$. This form for the PDE corresponds to the case of a random polycrystal, in which case C^0 is the volume average, or Voigt, effective modulus. If μ is taken in the form $\mu_{ijkl} = \mu(\mathbf{x}) \delta_{kj} \delta_{il}$ we can recover a simpler equation, for a scalar medium, that could correspond (in two dimensions) to an antiplane shear wave in a medium of constant density $\rho = 1$, and fluctuating shear modulus μ. In this case $\mathbf{G}$ represents antiplane displacement. The simplified version could also correspond to an acoustic pressure wave in a medium with constant bulk modulus, and fluctuating material density $1/\rho = \mu$. Neither of these interpretations of the similar version of Eq. (1) correspond to a practical system, but such interpretation may nevertheless be of utility for numerical simulations, as a scalar analog to the elastic wave equation for a medium with constant material density and fluctuating moduli, much like a polycrystal.

The perturbative technique to be employed considers effective properties to leading order in the strength of the fluctuations. As their mean $\langle \gamma \rangle$ is zero, the leading order non-trivial quantity is the moduli-covariance Λ, such that

$$\langle \gamma_{ijkl}(\mathbf{x}) \rangle = 0,$$

$$\langle \gamma_{ijkl}(\mathbf{x}) \gamma_{\alpha\beta\gamma\delta}(\mathbf{y}) \rangle = \Lambda(\mathbf{x} - \mathbf{y})^{\alpha\beta\gamma\delta}_{ijkl},$$
(3)

where the angular brackets $\langle \rangle$ are used to define the ensemble average quantities. On the assumption of statistical homogeneity, the moduli covariance has been taken to depend only on the difference between vector $\mathbf{x}$ and $\mathbf{y}$. In previous work we have taken the modulus covariance to factor into a scalar function of distance $|\mathbf{x} - \mathbf{y}|$ and a constant eighth rank tensor.[16,23]

The temporal Fourier transform of $\mathbf{G}$ is defined by

$$G_{i\alpha}(\mathbf{x}, \mathbf{y}, \omega) = \int_{-\infty}^{\infty} G_{i\alpha}(\mathbf{x}, \mathbf{y}, t) \exp(i\omega t) dt,$$
(4)

in which case Eq. (2) becomes

$$\left[(\omega + i\varepsilon)^2 \delta_{li} + C^0_{klij} \partial_k \partial_j + \frac{\partial}{\partial x_k} \gamma_{klij}(\mathbf{x}) \frac{\partial}{\partial x_j} \right] G_{i\alpha}(\mathbf{x}, \mathbf{y})$$

$$= \delta^3(\mathbf{x} - \mathbf{y}) \delta_{l\alpha},$$
(5)

where an $i\varepsilon$ has been included to emphasize the infinitesimal positive imaginary part of ω. Causality (the vanishing of G for negative values of t) assures that the Fourier transform is analytic in the upper half complex ω plane. Thus one understands ω to be complex: $\omega \to \omega + i\varepsilon$, where ε is an infinitesi-

mal positive quantity. This is a common device, and allows one to guarantee the existence of the Fourier transforms of nonsquare integrable functions, and to resolve the corresponding singularities.[16,44] Note that in Eq. (5) the ω dependence of $\mathbf{G}$ has been suppressed, but will be included from time to time for clarity.

The mean solution $\langle \mathbf{G} \rangle$ to Eq. (5) is expressible as the solution of an integral equation:

$$\langle G_{i\alpha}(\mathbf{x}, \mathbf{y}) \rangle = G^0_{i\alpha}(\mathbf{x}, \mathbf{y}) + \int \int G^0_{i\beta}(\mathbf{x}, \mathbf{z}) m_{\beta j}(\mathbf{z}, \mathbf{z}')$$

$$\times \langle G_{j\alpha}(\mathbf{z}', \mathbf{y}) \rangle d^3 z d^3 z',$$
(6)

in terms of the bare Green's function, $\mathbf{G}^0$ [$\mathbf{G}^0 \equiv$ the solution to Eq. (5) when $\gamma = 0$] and in terms of the self-energy operator $\mathbf{m}$. Equation (6) is termed a Dyson equation.[16,45]

The first-order smoothing approximation for the self-energy is given by

$$m_{\beta j}(\mathbf{z}, \mathbf{z}') = \left\langle \frac{\partial}{\partial z_\alpha} \gamma_{\alpha\beta\gamma\delta}(\mathbf{z}) \frac{\partial}{\partial z_\delta} G^0_{\gamma k}(\mathbf{z}, \mathbf{z}') \frac{\partial}{\partial z_i'} \gamma_{ikjl}(\mathbf{z}') \frac{\partial}{\partial z_1'} \right\rangle,$$
(7)

which is exact to this leading order in the modulus fluctuations (under the assumption that $\gamma \ll C^0$).

These expressions are simpler after employing a spatial Fourier transform. The spatial Fourier transform of an operator like $\mathbf{G}$ or $\mathbf{m}$, with two spatial arguments is defined by

$$\tilde{f}(\mathbf{q}, \mathbf{p}) = \frac{1}{(2\pi)^3} \int \int f(\mathbf{x}, \mathbf{y}) \exp[-i\mathbf{q} \cdot \mathbf{x} + i\mathbf{p} \cdot \mathbf{y}] d^3 x d^3 y.$$
(8)

Thus, the double spatial Fourier transform of $\mathbf{G}^0$ is

$$\tilde{G}^0_{i\alpha}(\mathbf{p}) \delta^3(\mathbf{p} - \mathbf{q}) = \frac{1}{(2\pi)^3} \int \int d^3 x d^3 y$$

$$\times \exp[+i\mathbf{q} \cdot \mathbf{x} - i\mathbf{p} \cdot \mathbf{y}] G^0_{i\alpha}(\mathbf{x}, \mathbf{y}).$$
(9)

The delta-function character is attributable to the homogeneity of the bare medium: $\mathbf{G}^0(\mathbf{x}, \mathbf{y}) = \mathbf{G}^0(\mathbf{x} - \mathbf{y})$. The Fourier transform in Eq. (9) has an inverse

$$G^0_{i\alpha}(\mathbf{x}, \mathbf{x}') = \frac{1}{(2\pi)^3} \int \int d^3 q d^3 p$$

$$\times \exp[i\mathbf{p} \cdot \mathbf{x} - i\mathbf{q} \cdot \mathbf{x}'] \tilde{G}^0_{i\alpha}(\mathbf{p}) \delta^3(\mathbf{p} - \mathbf{q}).$$
(10)

$\mathbf{G}^0(\mathbf{p})$ is readily constructed by Fourier transforming Eq. (5) for the case $\gamma = 0$ and then solving the resulting Christoffel equations

$$[(\omega + i\varepsilon)^2 - C^0_{klij} \mathbf{p}_k \mathbf{p}_j] \tilde{G}^0_{i\alpha}(\mathbf{p}) = \delta_{l\alpha},$$
(11)

or

$$\tilde{\mathbf{G}}^0(\mathbf{p}) = [\mathbf{I}(\omega + i\varepsilon)^2 - \mathbf{p} \cdot \mathbf{C}^0 \cdot \mathbf{p}]^{-1}.$$
(12)

Equation (12) may be inverted in closed form under some circumstances. If the Voigt moduli are isotropic, $\mathbf{G}^0$ may be decomposed into longitudinal and transverse parts,

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Ghoshal *et al.*: Diffuse backscatter in heterogeneous media   2011

$$G_{i\alpha}^0(\mathbf{p}) = \frac{1}{(\omega + i\varepsilon)^2 - c_L^2 p^2}\hat{p}_i\hat{p}_\alpha + \frac{1}{(\omega + i\varepsilon)^2 - c_T^2 p^2}$$

$$\times\{\delta_{i\alpha} - \hat{p}_i\hat{p}_\alpha\}. \tag{13}$$

The statistical homogeneity and isotropy of the medium also implies that the self-energy must have a double spatial Fourier transform with a delta-function form

$$-\sigma_{\beta j}(p)\delta^3(\mathbf{p} - \mathbf{q}) = \frac{1}{(2\pi)^3}\int\int d^3x\,d^3y$$

$$\times\exp\{-i\mathbf{p}\cdot\mathbf{x} + i\mathbf{q}\cdot\mathbf{y}\}m_{\beta j}(\mathbf{x},\mathbf{y}). \tag{14}$$

It also implies this form for the mean Green's function

$$\langle\widetilde{G}_{i\alpha}(\mathbf{p})\rangle\delta^3(\mathbf{p} - \mathbf{q}) = \frac{1}{(2\pi)^3}\int\int d^3x\,d^3y$$

$$\times\exp\{-i\mathbf{p}\cdot\mathbf{x} + i\mathbf{q}\cdot\mathbf{y}\}\langle G_{i\alpha}(\mathbf{x},\mathbf{y})\rangle. \tag{15}$$

The Dyson equation is easily solved in the spatial Fourier transform domain,

$$\langle\widetilde{G}_{i\alpha}(\mathbf{p})\rangle = \widetilde{G}_{i\alpha}^0(\mathbf{p}) - \widetilde{G}_{i\beta}^0(\mathbf{p})\sigma_{\beta k}(\mathbf{p})\langle\widetilde{G}_{k\alpha}(\mathbf{p})\rangle,$$

$$\langle\widetilde{\mathbf{G}}(\mathbf{p})\rangle = [[\widetilde{\mathbf{G}}^0(\mathbf{p})]^{-1} + \boldsymbol{\sigma}(\mathbf{p})]^{-1}$$

$$= [\mathbf{I}(\omega + i\varepsilon)^2 - \mathbf{p}\cdot\mathbf{C}^0\cdot\mathbf{p} + \boldsymbol{\sigma}(\mathbf{p})]^{-1}. \tag{16}$$

The effective wave number $\mathbf{p}$ is the wave number of the mean Green's function, the value of $\mathbf{p}$ at which $\langle\mathbf{G}\rangle$ is singular. These are the solutions $\mathbf{p}$ of the algebraic equation

$$\det[\mathbf{I}(\omega + i\varepsilon)^2 - \mathbf{p}\cdot\mathbf{C}^0\cdot\mathbf{p} + \boldsymbol{\sigma}(\mathbf{p})] = 0. \tag{17}$$

Thus, the problem of determining the mean response reduces to that of determining the self-energy. The real part of $\mathbf{p}$ is inversely proportional to the wave speed and the imaginary part is proportional to the attenuation coefficient α.

When the self-energy is given by the first-order smoothing approximation,[45] the Fourier transform of the self-energy (Eq. (7)) is

$$\sigma_{\beta j}(\mathbf{p}) = p_\alpha p_l\int G_{\gamma k}^0(\mathbf{r})\frac{\partial}{\partial r_\delta}\frac{\partial}{\partial r_i}[\exp\{-i\mathbf{p}\cdot\mathbf{r}\}\Lambda(\mathbf{r})_{ljki}^{\alpha\beta\gamma\delta}]d^3r, \tag{18}$$

where extensive use has been used made of integration by parts.

By defining the spatial Fourier transform of Λ as

$$\widetilde{\Lambda}(\mathbf{p})_{ijkl}^{\alpha\beta\gamma\delta} = \frac{1}{(2\pi)^3}\int\Lambda(\mathbf{r})_{ijkl}^{\alpha\beta\gamma\delta}\exp\{-i\mathbf{p}\cdot\mathbf{r}\}d^3r, \tag{19}$$

we can reexpress σ as a convolution in wave number space between $\mathbf{G}^0$ and Λ,

$$\sigma_{\beta j}(\mathbf{p}) = -p_\alpha p_l\int s_\delta s_i\widetilde{G}_{\gamma k}^0(\mathbf{s})\widetilde{\Lambda}_{ljki}^{\alpha\beta\gamma\delta}(\mathbf{p} - \mathbf{s})d^3s. \tag{20}$$

The Dyson equation governs the evolution of the mean field. Attenuations are then obtained from the imaginary part

of the effective wave number. The mean square response, the quantity needed to describe diffuse-field measurements, is proportional to the covariance of the mean Green's function. Therefore, the covariance of the mean Green's functions is now examined using the Bethe-Salpeter equation and ladder approximation.

III. MATHEMATICAL PRELIMINARIES II: BETHE-SALPETER EQUATION AND THE GREEN'S FUNCTION COVARIANCE

Calculations of mean responses, related to $\langle\mathbf{G}\rangle$, are not sufficient for the diffuse field measurements in which signals are squared before averaging. For this case, past work has examined the mean square Green's function. In general, however, the quantity of relevance is the Green's function covariance

$$\langle G_{\alpha\beta}(\mathbf{x},\mathbf{x}',\omega)G_{ij}^*(\mathbf{y},\mathbf{y}',\omega+\Omega)\rangle, \tag{21}$$

for which there also exists a multiple scattering formalism. We henceforth suppress the ω and $\omega+\Omega$ dependencies, taking the asterisk to imply $\omega+\Omega$ as well as complex conjugate. $\langle\mathbf{GG}^*\rangle$ may be inverse spatially Fourier transformed, with respect to all four spatial variables

$$\begin{smallmatrix}\mathbf{p}\alpha\\\mathbf{p}'i\end{smallmatrix}\mathbf{H}_{jq'}^{\beta\mathbf{q}}\delta^3(\mathbf{p} + \mathbf{q}' - \mathbf{q} - \mathbf{p}')$$

$$= \frac{1}{(2\pi)^6}\int d^3x\,d^3x'\,d^3y\,d^3y'\langle G_{\alpha\beta}(\mathbf{x},\mathbf{x}')G_{ij}^*(\mathbf{y},\mathbf{y}')\rangle$$

$$\times\exp\{-i\mathbf{px} + i\mathbf{qx}' + i\mathbf{p}'\mathbf{y} - i\mathbf{q}'\mathbf{y}'\}, \tag{22}$$

where the delta function is a consequence of the statistical homogeneity of the medium. It is conventional to employ the delta function and a change of variables $\mathbf{p}' = \mathbf{p} + \boldsymbol{\Delta}$, and consider only the three wave vector dependent quantity $\begin{smallmatrix}\mathbf{q}\\\mathbf{q}+\boldsymbol{\Delta}\end{smallmatrix}\mathbf{H}_{\mathbf{p}+\boldsymbol{\Delta}}^{\mathbf{p}}$.

The covariance is given by the solution of the Bethe-Salpeter equation[16,45]

$$\begin{smallmatrix}\mathbf{p}\\\mathbf{p}+\boldsymbol{\Delta}\,i\end{smallmatrix}{}^\alpha H_j^\beta\,{}_{\mathbf{q}+\boldsymbol{\Delta}}^{\mathbf{q}} = {}_i^\alpha\boldsymbol{\Gamma}_j^\beta\,{}_{\mathbf{q}+\boldsymbol{\Delta}}^{\mathbf{q}}\delta^3(\mathbf{p} - \mathbf{q})$$

$$+ \int d^3s\,{}_i^\alpha\Gamma_k^\gamma\,{}_{\mathbf{p}+\boldsymbol{\Delta}}^{\mathbf{p}}\,{}_{\mathbf{p}+\boldsymbol{\Delta}k}^{\mathbf{p}}{}^\gamma K_l^{\delta\,\mathbf{s}}\,{}_{\mathbf{s}+\boldsymbol{\Delta}}^{\mathbf{s}}\,{}_{\mathbf{s}+\boldsymbol{\Delta}l}^{\mathbf{s}}{}^\delta H_j^\beta\,{}_{\mathbf{q}+\boldsymbol{\Delta}}^{\mathbf{q}}, \tag{23}$$

where the double mean field Green's function Γ is

$${}_i^\alpha\Gamma_k^\gamma\,{}_{\mathbf{q}+\boldsymbol{\Delta}}^{\mathbf{q}} = \langle G_{\alpha\gamma}(\mathbf{q})\rangle\langle G_{ik}^*(\mathbf{q}+\boldsymbol{\Delta})\rangle, \tag{24}$$

which describes the propagation of the square of the mean field. The Bethe-Salpeter equation is formally exact, but approximation is necessary for the operator $\mathbf{K}$ (also known as the intensity operator). In what is often called the ladder approximation because of the shapes of the corresponding Feynman diagrams, $\mathbf{K}$ is approximated, like the mass operator, to leading order in the fluctuating moduli, i.e., to order Λ,

$$\begin{smallmatrix}\mathbf{p}\\\mathbf{p}+\boldsymbol{\Delta}k\end{smallmatrix}{}^\gamma K_l^{\delta\,\mathbf{s}}\,{}_{\mathbf{s}+\boldsymbol{\Delta}}^{\mathbf{s}} = p_\beta s_\alpha(p_i + \Delta_i)(s_j + \Delta_j)\widetilde{\Lambda}(\mathbf{p} - \mathbf{s})_{kijl}^{\gamma\beta\alpha\delta}. \tag{25}$$

The Bethe-Salpeter equation may be expanded in a multiple scattering series. To second order in $\mathbf{K}$ it is

$$\substack{\mathbf{p} \\ \mathbf{p}+\Delta} {}_i^\alpha H_j^\beta \substack{\mathbf{q} \\ \mathbf{q}+\Delta} = {}_i^\alpha \Gamma_j^\beta \substack{\mathbf{q} \\ \mathbf{q}+\Delta} \delta^3(\mathbf{p}-\mathbf{q})$$

$$+ {}_i^\alpha \Gamma_k^\gamma \substack{\mathbf{p} \\ \mathbf{p}+\Delta} \substack{\mathbf{p} \\ \mathbf{p}+\Delta k} {}^\gamma K_l^\delta \substack{\mathbf{q} \\ \mathbf{q}+\Delta l} {}^\delta \Gamma_j^\beta \substack{\mathbf{q} \\ \mathbf{q}+\Delta}$$

$$+ \int d^3 s {}_i^\alpha \Gamma_k^\gamma \substack{\mathbf{p} \\ \mathbf{p}+\Delta} \substack{\mathbf{p} \\ \mathbf{p}+\Delta k} {}^\gamma K_l^\delta \substack{\mathbf{s} \\ \mathbf{s}+\Delta}$$

$$\times {}_l^\delta \Gamma_m^\tau \substack{\mathbf{s} \\ \mathbf{s}+\Delta} \substack{\mathbf{s} \\ \mathbf{s}+\Delta m} {}^\tau K_l^\delta \substack{\mathbf{q} \\ \mathbf{q}+\Delta l} {}^\delta \Gamma_j^\beta \substack{\mathbf{q} \\ \mathbf{q}+\Delta}. \qquad (26)$$

The first term on the right-hand side of Eq. (26) describes a coherent propagation of covariance from source to receiver. The second term describes a coherent propagation at wave vector $\mathbf{q}$, followed by a scattering to wave vector $\mathbf{p}$. The third term in Eq. (26) is composed of five factors. The Γ factor on the right describes a coherent propagation, at wave vector $\mathbf{q}$, from the source to the first scattering, the next factor $\mathbf{K}$ describes a scattering to wave vector $\mathbf{s}$, another factor Γ describes coherent propagation at wave vector $\mathbf{s}$ to the second scatterer, at which the final wave vector $\mathbf{p}$ is produced. The final factor Γ, at wave vector $\mathbf{p}$, propagates the covariance coherently to the receiver. These expressions simplify substantially when Feynman diagrams are employed.

The above-mentioned mathematical preliminaries recapitulate, with minor generalizations, some of the material by Weaver.[16] In that work these equations were applied to the particular case of a point source and a point receiver in a statistically isotropic polycrystal composed of cubic crystallites. A diffusion equation was derived governing the elastic wave energy on sufficiently long time scales, longer than typical times between successive scatterings. An expression for the diffusivity was obtained in terms of microstructural parameters. Later, Turner and Weaver[21] applied these equations to a less severe limit in which scattering is assumed weak over distances of a wavelength, but in which time scales were not restricted. They transformed the Bethe-Salpeter equation into a radiative transfer equation similar to Eq. (1). Little attention was paid in that derivation to the roles played by particular sources and receivers, so the ambiguities mentioned in Sec. I were not addressed. Indeed, the purpose there was to derive rationally the otherwise heuristically asserted RTE and thereby to determine expressions for the scattering coefficients in that RTE. In this communication we propose to apply Eqs. (2), (3), and (7) to the particular case of a phase coherent source and receiver, and to determine a radiative-transfer-like equation governing the mean square signal produced at the receiver. This question is the more appropriate one if the intent is to model typical ultrasonic measurements.

IV. SOURCES, RECEIVERS, MEAN, AND MEAN SQUARE SIGNALS

We consider a medium with a body force that is distributed in space, time, and direction $B_\alpha(\mathbf{x})S(t)$. The factoring into distinct functions of space and time is characteristic of a real transducer, but is not required for subsequent analysis. The source produces a field

$$\Psi_\beta^S(\mathbf{x},t) = \int G_{\beta\alpha}(\mathbf{x},\mathbf{x}',t)B_\alpha(\mathbf{x}') \otimes S(t)d^3x', \qquad (27)$$

written here as a convolution in space and time between the Green's function of the medium and the source, where $\otimes$ represents a temporal convolution. The ensemble average of this response is

$$\langle \Psi_\beta^S(\mathbf{x},t)\rangle = \int \langle G_{\beta\alpha}(\mathbf{x},\mathbf{x}',t)\rangle B_\alpha(\mathbf{x}') \otimes S(t)d^3x', \qquad (28)$$

because the source distribution is nonstochastic.

The spatial and temporal Fourier transform of Eq. (28) is

$$\langle \Psi_\alpha^S(\mathbf{p},\omega)\rangle = \int \frac{d^3x\,dt}{(2\pi)^{3/2}}\langle \Psi_\alpha^S(\mathbf{x},t)\rangle \exp\{-i\mathbf{p}\cdot\mathbf{x}+i\omega t\}$$

$$= \int \frac{d^3x\,d^3x'\,dt}{(2\pi)^{3/2}}\langle G_{\alpha\beta}(\mathbf{x}-\mathbf{x}',t)\rangle B_\beta(\mathbf{x}')$$

$$\otimes S(t)\exp\{-i\mathbf{p}\cdot\mathbf{x}+i\omega t\}$$

$$= \int d^3r\langle G_{\alpha\beta}(\mathbf{r},t)\rangle \exp\{-i\mathbf{p}\cdot\mathbf{r}\}$$

$$\times \frac{d^3x'}{(2\pi)^{3/2}}B_\beta(\mathbf{x}')\exp\{-i\mathbf{p}\cdot\mathbf{x}\}$$

$$\otimes S(t)\exp\{i\omega t\}dt$$

$$= \langle \widetilde{G}_{\alpha\beta}(\mathbf{p},\omega)\rangle \widetilde{B}_\beta(\mathbf{p})S(\omega). \qquad (29)$$

Thus the Fourier transformed average field is given by a simple product of the Fourier transform of the source function and the average Green's function.

One can also define a receiver sensitivity distribution function $A_\alpha(\mathbf{x})R(t)$ which when convolved with a field Ψ gives the signal produced by the transducer. If convolved with the Green's function of the medium, it gives the response of the receiver at time t to a point source acting in direction β at time zero and position x is

$$\Psi_\beta^R(\mathbf{x},t) = \int R(t) \otimes A_\alpha(\mathbf{x}')G_{\alpha\beta}(\mathbf{x}',\mathbf{x},t)d^3\mathbf{x}'. \qquad (30)$$

Equation (30) can also be averaged and Fourier transformed giving

$$\langle \Psi_\beta^R(\mathbf{p},\omega)\rangle = \int \frac{d^3x\,dt}{(2\pi)^{3/2}}\langle \Psi_\beta^R(\mathbf{x},t)\rangle \exp\{i\mathbf{p}\cdot\mathbf{x}+i\omega t\}$$

$$= \langle \widetilde{G}_{\beta\alpha}(\mathbf{p},\omega)\rangle \widetilde{A}_\alpha(\mathbf{p})R(\omega). \qquad (31)$$

Note that the spatial Fourier transforms of Ψ^S and Ψ^R are defined with the opposite sign in the exponent.

If the field produced by the source is detected by the receiver, the resulting signal, $\phi(t)$, is given by the following convolution:

$$\phi(t) = \int\int R(t) \otimes \mathbf{A}_\beta(\mathbf{x})G_{\beta\alpha}(\mathbf{x},\mathbf{x}',t)B_\alpha(\mathbf{x}')$$

$$\otimes S(t)d^3x\,d^3x'. \qquad (32)$$

The position of the receiver and source is implicit in the factors $A_\beta(\mathbf{x})$ and $B_\alpha(\mathbf{x}')$, which will have their chief support

in the vicinity of those respective transducers. It is this expression that will serve formally to define the "signal," and of which we will take the mean and mean square. As the source and receiver characteristics are nonstochastic, the mean signal is obtained from Eq. (32) simply by replacing $\mathbf{G}$ with $\langle \mathbf{G} \rangle$. The mean square is more complicated, but clearly entails the covariance.

The temporal Fourier transform of $\langle \phi \rangle$ is

$$\langle \phi(\omega) \rangle = \int R(\omega) A_\alpha(\mathbf{x}) \langle G_{\alpha\beta}(\mathbf{x} - \mathbf{x}', \omega) \rangle$$
$$\times B_\beta(\mathbf{x}') S(\omega) d^3 x d^3 x'. \tag{33}$$

On substituting spatial Fourier representations for A and B one obtains

$$\langle \phi(\omega) \rangle = \int R(\omega) \widetilde{A}_\alpha(\mathbf{p}) \langle G_{\alpha\beta}(\mathbf{p}, \omega) \rangle \widetilde{B}_\beta(\mathbf{p}) S(\omega) d^3 p. \tag{34}$$

Diffuse field techniques typically analyze the square of the signal. Thus we need an expression for $\langle \phi^2(t) \rangle = \Phi(t)$ and its temporal Fourier transform given by

$$\Phi(\Omega) = \int_0^\infty \exp\{-i\Omega t\} \langle \phi^2(t) \rangle dt$$
$$= \int_0^\infty \exp\{-i\Omega t\} \langle \phi(t) \phi^*(t) \rangle dt$$
$$= \int_0^\infty \exp\{-i\Omega t\} \int_{-\infty}^\infty \int_{-\infty}^\infty \frac{d\omega}{2\pi} \frac{d\omega'}{2\pi}$$
$$\times \exp\{-i\omega t\} \exp\{i\omega' t\} \langle \phi(\omega) \phi^*(\omega') \rangle dt. \tag{35}$$

After performing the t and ω' integrations, it may be reexpressed in terms of the receiver and source characteristics and Green's function covariance

$$\Phi(\Omega) = \int d^3 x d^3 x' d^3 y d^3 y' \frac{d\omega}{2\pi} R(\omega) S(\omega) R^*(\omega + \Omega)$$
$$\times S^*(\omega + \Omega) A_\alpha(\mathbf{x}) A_i(\mathbf{y}) B_\beta(\mathbf{x}') B_j(\mathbf{y}')$$
$$\times \langle G_{\alpha\beta}(\mathbf{x}, \mathbf{x}') G_{ij}^*(\mathbf{y}, \mathbf{y}') \rangle, \tag{36}$$

where the covariance of the Green's function can be expressed in terms of its (12-fold) spatial inverse Fourier transform

$$\langle G_{\alpha\beta}(\mathbf{x}, \mathbf{x}') G_{ij}^*(\mathbf{y}, \mathbf{y}') \rangle$$
$$= \frac{1}{(2\pi)^6} \int d^3 p d^3 p' d^3 q d^3 q' \, {}_{\mathbf{p}'i}^{\mathbf{p}\alpha} \mathbf{H}_{jq'}^{\beta\mathbf{q}} \delta^3(\mathbf{p} + \mathbf{q}' - \mathbf{q} - \mathbf{p}')$$
$$\times \exp\{+i\mathbf{p} \cdot \mathbf{x} - i\mathbf{q} \cdot \mathbf{x}' - i\mathbf{p}' \cdot \mathbf{y} + i\mathbf{q}' \cdot \mathbf{y}'\}. \tag{37}$$

On substituting Eq. (37) into Eq. (36), and performing the integration over space, one obtains

$$\Phi(\Omega) = \int \frac{d\omega}{2\pi} d^3 p d^3 q d^3 \Delta \widetilde{A}_\alpha(\mathbf{p}) \widetilde{A}_i^*(\mathbf{p} + \boldsymbol{\Delta})$$
$$\times \widetilde{B}_\beta(\mathbf{q}) \widetilde{B}_j^*(\mathbf{q} + \boldsymbol{\Delta}) \, {}_{\mathbf{p}+\boldsymbol{\Delta}i}^{\mathbf{p}\quad\alpha} H_j^{\beta\quad\mathbf{q}}_{\ \ \mathbf{q}+\boldsymbol{\Delta}}$$
$$\times R(\omega) R^*(\omega + \Omega) S(\omega) S^*(\omega + \Omega). \tag{38}$$

This expression characterizes the mean square signal (in the frequency domain) in terms of the transducer properties and the scattering medium.

V. SCATTERED RESPONSE

We now substitute the multiple scattering series of Eq. (26) into Eq. (38) for Φ and conclude with a multiple scattering series expression for the mean square signal. The zeroth-order term is obtained most easily as

$$\Phi^0(\Omega) = \int \frac{d\omega}{2\pi} d^3 p d^3 p' \widetilde{A}_\alpha(\mathbf{p}) \widetilde{A}_i^*(\mathbf{p}') \widetilde{B}_\beta(\mathbf{p}) \widetilde{B}_j^*(\mathbf{p}') {}_i^\alpha \Gamma_j^\beta {}_{\mathbf{p}'}^{\mathbf{p}}$$
$$\times R(\omega) R^*(\omega + \Omega) S(\omega) S^*(\omega + \Omega)$$
$$= \left[\int \frac{d\omega}{2\pi} [d^3 p \widetilde{A}_\alpha(\mathbf{p}) \widetilde{B}_\beta(\mathbf{p}) R(\omega) S(\omega) \langle \widetilde{G}_{\alpha\beta}(\mathbf{p}, \omega) \rangle] \right.$$
$$\times \left[\int d^3 p' \widetilde{A}_i(\mathbf{p}') \widetilde{B}_j(\mathbf{p}') R(\omega + \Omega) S(\omega + \Omega) \right.$$
$$\left. \times \langle \widetilde{G}_{ij}(\mathbf{p}', \omega + \Omega) \rangle \right]^* \Bigg]$$
$$= \int \frac{d\omega}{2\pi} \langle \phi(\omega) \rangle \langle \phi(\omega + \Omega) \rangle^*, \tag{39}$$

which implies that

$$\Phi^0(T) = |\langle \phi(T) \rangle|^2. \tag{40}$$

To zeroth order the mean of the square is the square of the mean. A good diffuse field measurement is designed so that the mean response $\langle \phi \rangle$ vanishes. Consequently Φ^0 vanishes also.

To first order in powers of $\mathbf{K}$, $\Phi(\omega)$ is

$$\Phi^1(\Omega) = \int \frac{d\omega}{2\pi} d^3 p d^3 q d^3 \Delta [\widetilde{A}_\alpha(\mathbf{p}) \langle \widetilde{G}_{\alpha\beta}(p) \rangle R(\omega)]$$
$$\times [\widetilde{A}_i(\mathbf{p} + \boldsymbol{\Delta}) \langle \widetilde{G}_{ij}(\mathbf{p} + \boldsymbol{\Delta}) \rangle R(\omega + \Omega)]^*$$
$$\times {}_{\mathbf{p}+\boldsymbol{\Delta}j}^{\mathbf{p}\quad\beta} K_k^\gamma {}_{\mathbf{q}+\boldsymbol{\Delta}}^{\mathbf{q}} [\langle \widetilde{G}_{\gamma\tau}(\mathbf{q}) \rangle \widetilde{B}_\tau(p) S(\omega)]$$
$$\times [\langle \widetilde{G}_{kl}(\mathbf{q} + \boldsymbol{\Delta}) \rangle \widetilde{B}_l(p + \boldsymbol{\Delta}) S(\omega + \Omega)]^*. \tag{41}$$

The quantities in square brackets [ ] are the Fourier transforms of the (real-valued) fields Ψ^R and Ψ^S. Therefore,

$$\Phi^1(\Omega) = \int \frac{d\omega}{2\pi} d^3 p d^3 q d^3 \Delta \, {}_{\mathbf{p}+\boldsymbol{\Delta}j}^{\mathbf{p}\quad\beta} K_k^\gamma {}_{\mathbf{q}+\boldsymbol{\Delta}}^{\mathbf{q}}$$
$$\times \left[\int \langle \Psi_\beta^R(\mathbf{x}, t) \rangle \frac{dt d^3 x}{(2\pi)^{3/2}} \right.$$
$$\left. \times \exp\{i\mathbf{p} \cdot \mathbf{x} + i\omega t\} \right] \left[\int \langle \Psi_j^R(\mathbf{y}, u) \rangle \frac{du d^3 y}{(2\pi)^{3/2}} \right.$$
$$\left. \times \exp\{-i(\mathbf{p} + \boldsymbol{\Delta}) \cdot \mathbf{y} - i(\omega + \Omega)u\} \right]$$
$$\times \left[\int \langle \Psi_\gamma^S(\mathbf{x}', t) \rangle \frac{dt' d^3 x'}{(2\pi)^{3/2}} \exp\{-i\mathbf{q} \cdot \mathbf{x}' + i\omega t'\} \right]$$

$$\times \left[\int \langle \Psi_k^S(\mathbf{y}',u') \rangle \frac{du'd^3y'}{(2\pi)^{3/2}} \exp\{+ i(\mathbf{q}+\boldsymbol{\Delta})\cdot \mathbf{y}' \right.$$

$$\left. - i(\omega+\Omega)u'\} \right]. \tag{42}$$

On making further variable changes

$$\mathbf{x} = \mathbf{X} + \boldsymbol{\xi}/2, \quad \mathbf{y} = \mathbf{X} - \boldsymbol{\xi}/2, \quad t = T + \tau/2, \quad u = T - \tau/2,$$

$$\mathbf{x}' = \mathbf{X}' + \boldsymbol{\xi}'/2, \quad \mathbf{y}' = \mathbf{X}' - \boldsymbol{\xi}'/2, \quad t' = T' + \tau'/2, \tag{43}$$

$$u' = T' - \tau'/2,$$

we obtain

$$\Phi^1(\Omega) = \int \frac{d\omega}{(2\pi)^7} d^3p\, d^3q\, d^3\Delta \, {}_{\mathbf{p}+\boldsymbol{\Delta}j}^{\,\mathbf{p}\,\beta} K_k^{\gamma\,\mathbf{q}}{}_{\mathbf{q}+\boldsymbol{\Delta}}$$

$$\times \int d^3X\, dT\, d^3\xi\, d\tau \langle \psi_\beta^R(\mathbf{X}+\boldsymbol{\xi}/2, T+\tau/2) \rangle$$

$$\times \langle \Psi_j^R(\mathbf{X}-\boldsymbol{\xi}/2, T-\tau/2) \rangle \exp\{i\tau(\omega+\Omega/2)$$
$$+ i\boldsymbol{\xi}\cdot(\mathbf{p}+\boldsymbol{\Delta}/2) - i\Omega T - i\boldsymbol{\Delta}\cdot\mathbf{X}\}$$

$$\times \int d^3X'\, dT'\, d^3\xi'\, d\tau' \langle \Psi_\gamma^S(\mathbf{X}'+\boldsymbol{\xi}'/2, T'+\tau'/2) \rangle$$

$$\times \langle \Psi_k^S(\mathbf{X}'-\boldsymbol{\xi}'/2, T'-\tau'/2) \rangle \exp\{i\tau'(\omega+\Omega/2)$$
$$- i\boldsymbol{\xi}'\cdot(\mathbf{q}+\boldsymbol{\Delta}/2) - i\Omega T' + i\boldsymbol{\Delta}\cdot\mathbf{X}'\}. \tag{44}$$

The integrations over $\boldsymbol{\xi}$, τ, $\boldsymbol{\xi}'$, and τ' can be done by reference to a definition of a four-fold Wigner transform. For source and receiver fields we use opposite sign conventions,

$$W_{\gamma k}^S(\mathbf{X}, T, \mathbf{k}, \omega)$$

$$= \int \langle \Psi_\gamma^S(\mathbf{X}+\boldsymbol{\xi}/2, T+\tau/2) \rangle$$

$$\times \langle \Psi_k^S(\mathbf{X}-\boldsymbol{\xi}/2, T-\tau/2) \rangle \exp\{- i\mathbf{k}\cdot\boldsymbol{\xi} + i\omega\tau\} d^3\xi\, d\tau,$$

$$W_{\beta j}^R(\mathbf{X}, T, \mathbf{k}, \omega)$$

$$= \int \langle \Psi_\beta^R(\mathbf{X}+\boldsymbol{\xi}/2, T+\tau/2) \rangle$$

$$\times \langle \Psi_j^R(\mathbf{X}-\boldsymbol{\xi}/2, T-\tau/2) \rangle \exp\{+ i\mathbf{k}\cdot\boldsymbol{\xi} + i\omega\tau\} d^3\xi\, d\tau. \tag{45}$$

Equation (45) is spatial and temporal generalizations of the more well known temporal Wigner transform widely discussed in signal processing[42,46] and others areas such as imaging,[47] optics,[48–50] and so on. W represents a distribution in space and time $\mathbf{X}, T$ of spectral energy density as a function of wave vector $\mathbf{k}$ and frequency ω.

Using the definitions in Eq. (45) we conclude with an expression for the singly scattered contribution to Φ:

$$\Phi^1(\Omega) = \int \frac{d\omega}{(2\pi)^7} d^3p\, d^3q\, d^3\Delta \, {}_{\mathbf{p}+\boldsymbol{\Delta}j}^{\,\mathbf{p}\,\beta} K_k^{\gamma\,\mathbf{q}}{}_{\mathbf{q}+\boldsymbol{\Delta}}$$

$$\times \int d^3X\, dT\, W_{\beta j}^R(\mathbf{X}, T, \mathbf{p}+\boldsymbol{\Delta}/2, \omega+\Omega/2)$$

$$\times \exp\{- i\Omega T - i\boldsymbol{\Delta}\cdot\mathbf{X}\}$$

$$\times \int d^3X'\, dT'\, W_{\gamma k}^S(\mathbf{X}', T', \mathbf{q}+\boldsymbol{\Delta}/2, \omega+\Omega/2)$$

$$\times \exp\{- i\Omega T' + i\boldsymbol{\Delta}\cdot\mathbf{X}'\}. \tag{46}$$

At this point it is necessary to argue that the Δ and Ω dependence in $\mathbf{K}$ is negligible. This assumption is the key approximation that allows radiative transfer-like equations to be obtained from the otherwise more general Bethe-Salpeter equation. This then allows the Δ integration, and the integration over Ω entailed in the inverse Fourier transform needed to recover $\Phi(t)$, to be done immediately. One finds, after changing variables by means of a shift of ω by $\Omega/2$ and a shift of $\mathbf{p}$ and $\mathbf{q}$ by $\Delta/2$,

$$\Phi^1(T) = \int \frac{d\omega}{(2\pi)^4} d^3p\, d^3q\, d^3X\, dT'$$

$$\times W_{\beta j}^R(\mathbf{X}, T-T', \mathbf{p}, \omega) \, {}_{\mathbf{p}j}^{\,\beta} K_k^{\gamma\,\mathbf{q}}{}_{\mathbf{q}} W_{\gamma k}^S(\mathbf{X}, T', \mathbf{q}, \omega). \tag{47}$$

In the ladder approximation for $\mathbf{K}$ as shown in Eq. (25), we see that there is no Ω dependence. Media for which the inhomgeneities entail factors of ω (for example, inhomogeneities in mass density) will generate weak Ω dependence in $\mathbf{K}$; media for which the scatterers have dynamic internal variables (a case in which internal variables are resonant) will generate a stronger Ω dependence in $\mathbf{K}$, even with the ladder approximation. If one employs a more complex approximation for $\mathbf{K}$ than the ladder approximation, e.g., a $\mathbf{K}$ that includes processes by which there is an internal propagation by $\langle \mathbf{G} \rangle$, there will be an Ω dependence. All such dependencies correspond to time delays between the arrival of a coherent wave at a scatterer and its subsequent reemision from a scatterer. As we often take the scatterers to have no significant frequency dependence we do not further consider this possibility here.

$\mathbf{K}$ has, even with the ladder approximation, an explicit dependence on Δ. If we were to include the several terms involving Δ, then it would be augmented by extra terms, but with the derivatives of the $\mathbf{W}$'s with respect to $\mathbf{X}$ in lieu of the $\mathbf{W}$'s themselves. These additional terms would be smaller by factors of the ratio of the length scale over which $\mathbf{W}$ varies to the length scale given by $\mathbf{p}$ or $\mathbf{q}$. Thus as long as the Wigner function of the source and receiver fields varies, with $\mathbf{X}$, slowly on the scale of the wavelength, the Δ dependence in $\mathbf{K}$ is safely neglected. This assumption, that $\Delta \ll \omega/c_L$, is the same one that allows the Bethe-Salpeter equation to be recast as an equation of radiative transfer.[16]

This requirement does assert some interesting restraints on the application of Eq. (47). Because $\mathbf{W}$ must attenuate like the energy of a coherent field, it implies that the attenuation of the mean Green's function should be moderate or

small over distances of a wavelength: $\alpha\lambda \ll 2\pi$. It also implies that the near field of a point source (and the focal region of a sharply focused source) must be a minor part of the multiple scattering process, this is, that very little scattering takes place in any region where the field is changing amplitude over length scales of the order of $\lambda/2\pi$. Because the mean free path is equal to or somewhat longer than $1/2\alpha$, this constraint is also equivalent to $\alpha\lambda \ll 2\pi$.

From Eq. (47), we see that the singly scattered contribution is a convolution in time and space of the product of the Wigner distributions of the beam pattern of the source and receiver, mediated by the scattering strength **K** from wave vector **q** to wave vector **p**. The Wigner distributions that enter into this description are the Wigner transforms of the coherent fields associated with source and receiver, the fields that would be present if the source and receiver were placed in a medium described by the average Green's function. Equation (47) is one of the key results of this article.

VI. WIGNER TRANSFORM OF A PISTON TRANSDUCER

The variance of the signal obtained from a typical C-Scan is defined here as the singly scattered response. In such experiments the signals are acquired at various positions of the transducer and the scattering from the focal depth is analyzed to infer microstructural information by examining the statistics of the signals. The variance of the signal at the focal region may be modeled using Eq. (47). Thompson and Gray previously derived an expression for the singly scattered response (SSR) for polycrystalline materials.[51] A comparison of their theoretical model with experiment results was first presented by Margetan et al.[17] Their model was derived for longitudinal-to-longitudinal scattering only. Here, Eq. (47) is used to derive an expression similar to that of Thompson and Gray but within the full multiple-scattering context derived earlier. Such a derivation will allow the limits of the single-scattering assumption to be explored more fully in the future.

Equation (47) is examined for a specific case used extensively for microstructure measurements, for which the same transducer is used as both source and receiver. Since the focus here is on the longitudinal-to-longitudinal SSR, the Wigner transform of the receiver and the source simplify from the generalized tensorial form given by Eq. (45).

The longitudinal component of the mean Green's function in space and time is given by[16]

$$\langle \mathbf{G}(\mathbf{x},\mathbf{x}',t) \rangle = -\frac{\exp[-\alpha^L|\mathbf{x}-\mathbf{x}'|]}{4\pi c_L^2(|\mathbf{x}-\mathbf{x}'|)}\delta\left(t - \frac{|\mathbf{x}-\mathbf{x}'|}{c_L}\right)\hat{\mathbf{p}}\hat{\mathbf{p}},$$

(48)

where α^L is the longitudinal attenuation coefficient.

Here, the body force is assumed to be a Gaussian pulse in space at $z=0$ and is given by

$$B_\alpha(\mathbf{x}) = B_0\delta(z)\exp\left[-\frac{x^2+y^2}{w_0^2}\right]\hat{\mathbf{n}}_\alpha,$$

(49)

where w_0 is the effective transducer radius, B_0 is the force per unit area, and $\hat{\mathbf{n}}_\alpha$ is the unit normal to the transducer face. Thompson et al.[52] calculated w_0 and B_0 and explained the procedure to obtain the specific values for calibration of a transducer. The source is also assumed to be a Gaussian pulse in time such that

$$S(t) = \frac{1}{\sigma\sqrt{\pi}}\exp\left[i\omega_0 t - \frac{t^2}{\sigma^2}\right],$$

(50)

where ω_0 is the forcing frequency and σ is the width of the pulse. The mean source field is

$$\langle \Psi_\beta(\mathbf{x},t) \rangle = \int \langle G_{\beta\alpha}(\mathbf{x},\mathbf{x}',t)\rangle B_\alpha(\mathbf{x}') \otimes S(t) d^3x',$$

(51)

where $\otimes$ denotes a temporal convolution. The temporal convolution becomes

$$\langle G_{\beta\alpha}(\mathbf{x},\mathbf{x}',t)\rangle \otimes S(t) = -\frac{\exp\left(i\omega_0 t - \left(\frac{t}{\sigma}\right)^2\right)\exp\left[\left(-ik_0 + \frac{2t}{\sigma^2 c_L}\right)|\mathbf{x}-\mathbf{x}'| - \frac{|\mathbf{x}-\mathbf{x}'|^2}{\sigma^2 c_L^2}\right]}{\sigma\sqrt{\pi}4\pi c_L^2}\frac{}{(|\mathbf{x}-\mathbf{x}'|)}\exp[-\alpha^L|\mathbf{x}-\mathbf{x}'|]\hat{p}_\beta\hat{p}_\alpha,$$

(52)

where $|\mathbf{x}-\mathbf{x}'| = \sqrt{(z-z')^2+(x-x')^2+(z-z')^2}$, and $k_0 = \omega_0/c_L$. Within the context of the paraxial approximation, the square root is expanded to first order as $((z-z')+\{(x-x')^2+(y-y')^2\}/2z)$ and to zeroth order in the amplitude term as $(z-z')$. Such an approximation, although not necessary, allows Eq. (47) to be simplified greatly. Thus, under this approximation Eq. (52) becomes

$$\langle G_{\beta\alpha}(\mathbf{x},\mathbf{x}',t)\rangle \otimes S(t) = -\frac{1}{\sigma\sqrt{\pi}4\pi c_L^2(z-z')}\exp\left(i\omega_0 t - \left(\frac{t}{\sigma}\right)^2\right)\exp[-\alpha^L(z-z')]$$

$$\times\exp\left[-ik_0((z-z')+\{(x-x')^2+(y-y')^2\}/2z) + \frac{2t}{\sigma^2 c_L}(z-z') - \frac{(z-z')^2}{\sigma^2 c_L^2}\right]\hat{p}_\beta\hat{p}_\alpha.$$

(53)

Performing the convolution in space reduces Eq. (51) to

$$\langle \Psi_\beta(\mathbf{x},t)\rangle = -\frac{B_0 \exp\left(i\omega_0 t - \left(\frac{t}{\sigma}\right)^2\right)}{\sigma\sqrt{\pi}4\pi c_L^2}\exp\left[-\alpha^L z + \frac{z(2c_L t - z)}{\sigma^2 c_L^2}\right]\frac{\lambda_0 w_0}{w(z)}$$

$$\times \exp\left[-ik_0 z - \frac{r^2}{w^2(z)} - i\frac{\pi}{\lambda_0}\frac{r^2}{R(z)} + i\frac{\pi}{2} + i\Psi_0(z)\right]\hat{p}_\beta(\hat{\mathbf{p}}\cdot\hat{\mathbf{n}}). \tag{54}$$

In Eq. (54), $w(z)$ is the Gaussian beam width, $R(z)$ the radius of curvature of the wave front, and $\Psi_0(z)$ is the excess phase defined by[53]

$$w(z) = w_0\left[1 + \left(\frac{\lambda_0 z}{\pi w_0^2}\right)^2\right]^{1/2},$$

$$R(z) = z + \left(\frac{\pi w_0^2}{\lambda_0}\right)^2\frac{1}{z},$$

$$\Psi_0(z) = \tan^{-1}\left(\frac{\lambda_0 z}{\pi w_0^2}\right), \tag{55}$$

where $\lambda_0 = 2\pi/k_0$. The mean source field can also be written in the form

$$\langle \Psi_\beta^S(\mathbf{r},z,t)\rangle = A(\mathbf{r},z,t)e^{i\Theta(\mathbf{r},z,t)}\hat{p}_\beta(\hat{\mathbf{p}}\cdot\hat{\mathbf{n}}), \tag{56}$$

where $r^2 = x^2 + y^2$. The amplitude $A(\mathbf{r},z,t)$ and the phase $\Theta(\mathbf{r},z,t)$ of the source field are given as

$$A(\mathbf{r},z,t) = -\frac{B_0 \exp\left(-\left(\frac{t}{\sigma}\right)^2\right)}{\sigma\sqrt{\pi}4\pi c_L^2}\frac{\lambda_0 w_0}{w(z)}$$

$$\times\exp\left[-\frac{r^2}{w^2(z)} - \alpha^L z + \frac{z(2c_L t - z)}{\sigma^2 c_L^2}\right],$$

$$\Theta(\mathbf{r},z,t) = \omega_0 t - k_0 z - \frac{\pi}{\lambda_0}\frac{r^2}{R(z)} + \frac{\pi}{2} + \Psi_0(z). \tag{57}$$

Using high-frequency asymptotics, where it is assumed that $A(\mathbf{r},z,t)$ varies much slower than $e^{i\Theta(\mathbf{r},z,t)}$, the Wigner transform of $\langle \Psi_\beta^S\rangle$ can be approximated by[36,54]

$$W(\mathbf{x},t,\mathbf{k},\omega) = (2\pi)^4|A(\mathbf{x},t)|^2\delta^3(\mathbf{k} - \nabla_{\mathbf{x}}\Theta(\mathbf{x},t))$$

$$\times\delta\left(\omega - \frac{\partial}{\partial t}\Theta(\mathbf{x},t)\right). \tag{58}$$

Thus the space-time Wigner transform of the mean field becomes

$$W_{\gamma k}(\mathbf{x},\mathbf{k},t,\omega)$$

$$= (2\pi)^4\left(\frac{B_0}{\sigma\sqrt{\pi}4\pi c_L^2}\frac{\lambda_0 w_0}{w(z)}\right)^2$$

$$\times\exp\left[-2\alpha^L z + 2z\frac{(2c_L t - z)}{\sigma^2 c_L^2} - 2\frac{r^2}{w^2(z)} - 2\left(\frac{t}{\sigma}\right)^2\right]$$

$$\times\delta(\omega - \omega_0)\delta^2\left(\mathbf{k}_r + \frac{2\pi}{\lambda_0}\frac{r}{R(z)}\right)\delta\left(k_z + k_0 - \frac{\lambda_0}{\pi}\frac{1}{w^2(z)}\right)$$

$$-\frac{\pi}{\lambda_0}r^2\frac{1}{R^2(z)}\left(1 - \left(\frac{\pi w_0^2}{\lambda_0 z}\right)^2\right)\right)(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}})^2\hat{k}_\gamma\hat{k}_k. \tag{59}$$

The terms in the delta functions

$$\frac{2\pi}{\lambda_0}\frac{r}{R(z)}, \quad -\frac{\lambda_0}{\pi}\frac{1}{w^2(z)} - \frac{\pi}{\lambda_0}r^2\frac{1}{R^2(z)}\left(1 - \left(\frac{\pi w_0^2}{\lambda_0 z}\right)^2\right) \tag{60}$$

are assumed negligible when compared with k_0. This implies that all the energy is primarily in the $k_z = k_0$ direction, which is along the transducer axis ($|\mathbf{k}|^2 = k_r^2 + k_z^2 \cong k_z^2$). Therefore the sidelobes are neglected in what follows. Finally the longitudinal component of the Wigner transform of a piston transducer becomes

$$W_{\gamma k}^S(\mathbf{x},\mathbf{k},t,\omega)$$

$$= (2\pi)^4\left(\frac{B_0}{\sigma\sqrt{\pi}4\pi c_L^2}\frac{\lambda_0 w_0}{w(z)}\right)^2$$

$$\times\exp\left[-2\alpha^L z + 2z\frac{(2c_L t - z)}{\sigma^2 c_L^2} - 2\frac{r^2}{w^2(z)} - 2\left(\frac{t}{\sigma}\right)^2\right]$$

$$\times\delta(\omega - \omega_0)\delta^3(\mathbf{k} + k_0)(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}})^2\hat{k}_\gamma\hat{k}_k. \tag{61}$$

Equation (61) is the second primary result of this article. It describes the distribution of longitudinal energy in space, time, frequency, and wave vector resulting from a piston transducer. It should be noted that the delta function expressions for frequency and wave vector may correspond to limits of Gaussians, a result that depends on the ansatz for the Wigner transform (e.g., time-space, frequency-space, time-wave vector, frequency-wave vector).

VII. SINGLY SCATTERED RESPONSE FROM A PISTON TRANSDUCER

Using the expression for $\mathbf{W}$ given in Eq. (61) for both the source and receiver, Eq. (47) is now simplified. First the spatial and the tensorial components are assumed to be independent and written as[55]

$$\tilde{\Lambda}(\mathbf{p})_{lmjk}^{\alpha\beta\gamma\delta} = \Xi_{lmjk}^{\alpha\beta\gamma\delta}\eta(\mathbf{p}), \tag{62}$$

where $\tilde{\eta}(\mathbf{p})$ is the Fourier transform of the spatial correlation function. Next the intensity operator is written from the definition in Eq. (25) as (neglecting the slowly varying Δ dependence)

$$_{pj}^{p\beta}K_{kq}^{\gamma q} \approx \eta(\mathbf{p} - \mathbf{q})\mathbf{p}_\alpha\mathbf{q}_\delta\mathbf{p}_l\mathbf{q}_m\Xi_{lmjk}^{\alpha\delta\beta\gamma}$$

$$= p^2q^2\eta(p\hat{\mathbf{p}} - q\hat{\mathbf{q}})\hat{p}_\alpha\hat{q}_\delta\hat{p}_l\hat{q}_m\Xi_{lmjk}^{\alpha\delta\beta\gamma}. \tag{63}$$

The SSR then becomes

$$\Phi^1(t) = (2\pi)^8 \left(\frac{B_0\lambda_0}{\sigma\sqrt{\pi}4\pi c_L^2}\right)^4 \int \frac{dw}{(2\pi)^6}\left[\left(\frac{\omega_0}{c_L}\right)^2 \eta\left(\frac{\omega_0}{c_L}\hat{\mathbf{p}} - \frac{\omega_0}{c_L}\hat{\mathbf{q}}\right)\right.$$

$$\times \hat{p}_\alpha\hat{q}_\delta\hat{p}_l\hat{q}_m \Xi_{lmjk}^{\alpha\delta\beta\gamma}\left]\left(\frac{w_0}{w(z)}\right)^4\right.$$

$$\times \exp\left[-4\alpha^L z + 2z\frac{2c_L(t-t')-z}{\sigma^2 c_L^2} + 2z\frac{2c_L t'-z}{\sigma^2 c_L^2}\right.$$

$$\left. -4\frac{r^2}{w^2(z)} - 2\frac{(t-t')^2}{\sigma^2} - 2\frac{t'^2}{\sigma^2}\right]$$

$$\times \delta^2(\hat{\mathbf{p}}+\hat{\mathbf{p}}_0)\hat{p}_\beta\hat{p}_j(\hat{\mathbf{p}}\cdot\hat{\mathbf{n}})^2\delta(\omega-\omega_0)$$

$$\times \delta^2(\hat{\mathbf{q}}+\hat{\mathbf{q}}_0)\hat{q}_\gamma\hat{q}_k(\hat{\mathbf{q}}\cdot\hat{\mathbf{n}})^2\delta(\omega-\omega_0)d^2\hat{p}d^2\hat{q}d^3Xdt'. \tag{64}$$

The integration over $\hat{\mathbf{p}}$, $\hat{\mathbf{q}}$, and ω are trivial due to the δ function dependence. In typical diffuse backscatter measurements the angle between the propagation and scattered direction is π for a normal incidence pulse echo setup. Therefore, $\hat{\mathbf{p}}_0\cdot\hat{\mathbf{n}}=1$, $\hat{\mathbf{q}}_0\cdot\hat{\mathbf{n}}=-1$, and $\hat{\mathbf{p}}_0\cdot\hat{\mathbf{q}}_0=-1$. Thus the SSR is

$$\Phi^1(t) = (2\pi)^2\left(\frac{B_0\lambda_0}{\sigma\sqrt{\pi}4\pi c_L^2}\right)^4 \int \left(\frac{\omega_0}{c_L}\right)^2 \eta(\theta_{p_0q_0}=\pi)$$

$$\times [\hat{p}_{0\beta}\hat{p}_{0j}\hat{p}_{0\alpha}\hat{q}_{0\delta}\hat{p}_{0l}\hat{q}_{0m}\hat{q}_{0\gamma}\hat{q}_{0k}\Xi_{lmjk}^{\alpha\delta\beta\gamma}]\left(\frac{\omega_0}{w(z)}\right)^4$$

$$\times \exp\left[-4\alpha^L z + 4z\frac{c_L t-z}{\sigma^2 c_L^2} - 4\frac{r^2}{w^2(z)}\right.$$

$$\left. -2\frac{(t-t')^2}{\sigma^2} - 2\frac{t'^2}{\sigma^2}\right]d^3Xdt', \tag{65}$$

where $\eta(\theta_{p_0s_0})=\eta[(\omega/c_L)\hat{\mathbf{p}}_0-(\omega/c_L)\hat{\mathbf{q}}_0]$ and $\theta_{p_0q_0}$ is the angle between the vector $\hat{\mathbf{p}}_0$ and $\hat{\mathbf{q}}_0$ [$\hat{\mathbf{p}}_0\cdot\hat{\mathbf{q}}_0=\cos(\theta_{p_0q_0})$]. Let the inner product be denoted in direct form by

$$\Xi_{lmjk}^{\alpha\delta\beta\gamma}\hat{p}_\alpha\hat{p}_l\hat{q}_\delta\hat{q}_m\hat{p}_\beta\hat{p}_j\hat{q}_\gamma\hat{q}_k = \Xi_{\cdots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{q}}\hat{\mathbf{q}}}^{\cdots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{q}}\hat{\mathbf{q}}}(\theta_{pq}). \tag{66}$$

The singly scattered response is then

$$\Phi^1(t) = (2\pi)^2\left(\frac{B_0\lambda_0}{\sigma\sqrt{\pi}4\pi c_L^2}\right)^4\left[\left(\frac{\omega_0}{c_L}\right)^4 \eta(\pi)\Xi_{\cdots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{q}}\hat{\mathbf{q}}}^{\cdots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{q}}\hat{\mathbf{q}}}(\pi)\right]$$

$$\times \int\left(\frac{w_0}{w(z)}\right)^4\exp\left[-4\alpha^L z + 4z\frac{c_L t-z}{\sigma^2 c_L^2} - 4\frac{r^2}{w^2(z)}\right.$$

$$\left. -2\frac{(t-t')^2}{\sigma^2} - 2\frac{t'^2}{\sigma^2}\right]d^3Xdt'. \tag{67}$$

The integration in the lateral dimensions, x and y, characterizes the effect of scattering from a plane at depth z in the material.[17] It is given by

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left[-4\frac{x^2+y^2}{w^2(z)}\right]dxdy = \frac{\pi w^2(z)}{4}. \tag{68}$$

The temporal integration becomes

$$\int_{-\infty}^{\infty}\exp\left(-2\frac{(t-t')^2+(t')^2}{\sigma^2}\right)dt' = \frac{\sigma\sqrt{\pi}}{2}\exp\left[-\frac{t^2}{\sigma^2}\right]. \tag{69}$$

Substituting Eqs. (68) and (69) into Eq. (67), the SSR becomes

$$\Phi^1(t) = \Phi_0\frac{1}{w_0^2}\left[\frac{\pi}{2}\frac{\omega_0^4}{c_L^8}\eta(\pi)\Xi_{\cdots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{q}}\hat{\mathbf{q}}}^{\cdots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{q}}\hat{\mathbf{q}}}(\pi)\right]\exp\left(-\frac{t^2}{\sigma^2}\right)$$

$$\times \int_{-\infty}^{\infty}\left(\frac{\pi}{4}\frac{w_0^4}{w^2(z)}\right)\exp\left(-4\alpha^L z + 4z\frac{c_L t-z}{\sigma^2 c_L^2}\right)dz, \tag{70}$$

where

$$\Phi_0 = \frac{B_0^4 w_0^2}{4\sqrt{\pi}\sigma^3\omega_0^4}. \tag{71}$$

Equation (71) is defined such that $\Phi^1(t)/\Phi_0$ is a dimensionless quantity. The parameters σ, w_0, and B_0 in Eq. (71) may be obtained while calibrating and characterizing transducers for typical ultrasonic pulse-echo experiments (i.e., B_0 can be calculated from the amplitude of the reflected signal from the surface of the test sample). Also, σ and w_0 are related to a Gaussian envelope fit to the input signal.

The integrand in Eq. (70) is related to the transducer beam model. The term in square brackets $[(\pi/2)\times(\omega_0^4/c_L^8)\eta(\pi)\Xi_{\cdots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{q}}\hat{\mathbf{q}}}^{\cdots\hat{\mathbf{p}}\hat{\mathbf{p}}\hat{\mathbf{q}}\hat{\mathbf{q}}}(\pi)]$ is known as the diffuse backscatter coefficient,[17,56] which depends on the microstructural properties of the material and thus controls the amplitude of the SSR. Equation (70) is the third primary result of this article, a quantity that may be compared directly with experiments involving transducers. In a typical diffuse backscatter experiment the standard deviation of the signals [square root of Eq. (70)] is often used.[17]

In Sec. VIII the SSR model, given by Eq. (70), will be shown to match qualitatively with the Thompson and Gray model.[51] Their model was developed to include the interface between a liquid and solid boundary, an aspect that is not currently included in the derivation presented here. A rigorous comparison between the two models is now under development. However, it should be noted that differences between the two models are to be expected since the two models are fundamentally different, with the present model based on the fourfold Wigner transform of the transducer beam pattern.

VIII. EXAMPLE RESULTS

Example results are now presented for specific cases of polycrystalline aluminum and iron. The single-crystal material properties used for aluminum are $c_{11}=103.4$ GPa, $c_{12}=57.1$ GPa, $c_{44}=28.6$ GPa, $\rho=2700$ kg/m^3 ($c_L=6289$ m/s) and for iron are $c_{11}=216$ GPa, $c_{12}=145$ GPa, $c_{44}=129$ GPa, $\rho=8000$ kg/m^3 ($c_L=5740$ m/s). The ultrasonic beam from a focused transducer converges to a plane known as the focal plane. The best resolution is generally obtained at the focal depth. Hence, the peak amplitude of the scattered response appears near the focal plane in the material. In the following

examples, the transducer diameter is chosen as $w_0 = 6.35$ mm and the focus into the material is 12 mm.

The SSR given by Eq. (70) for a piston transducer can be easily used for a focused Gaussian beam using the following definitions for the beam width and the curvature:[52]

$$q(z) = q(0) + z,$$

$$\frac{1}{q(z)} = \frac{1}{R(z)} - \frac{i\lambda_0}{\pi w^2(z)},$$

$$R(z) = \frac{1}{\text{Re}[1/q(z)]}, \quad w(z) = \left[\frac{-\lambda_0/\pi}{\text{Im}[1/q(z)]} \right]^{1/2}, \quad (72)$$

where $R(0) = -F$ with F the focal point. These transformations do not violate any assumptions made previously as long as the focus is not near the transducer surface. Equation (72) reduces to Eq. (55) as $R(0) \rightarrow \infty$ (for a planar transducer the focal point approaches infinity).

The Fourier transform of the two-point correlation function $\eta(r) = e^{-r/L}$ at $\theta_{pq} = \pi$ is given by

$$\eta(k) = \frac{L^3}{\pi^2 (1 + k^2 L^2)^2},$$

where L is the correlation length. A correlation length of 20 μm is used for both materials (for aluminum $\alpha^L = 0.47$ Np/m; for iron $\alpha^L = 8.5$ Np/m at 15 MHz). Weaver[16] has given the inner product for scattering from cubic grains which could also be generalized to other material symmetries. The inner product for cubic symmetry is

$$\Xi_{\cdots\hat{p}\hat{p}\hat{q}\hat{q}}^{\cdots\hat{p}\hat{p}\hat{q}\hat{q}}(\theta_{pq}) = \frac{\nu^2}{\rho^2} \left[\frac{9}{525} + \frac{6}{526} \cos^2 \theta_{pq} + \frac{1}{525} \cos^4 \theta_{pq} \right], \quad (73)$$

where ρ is the density and $\nu = c_{11} - c_{12} - 2c_{44}$ for cubic symmetry.

The first example result is that of the Wigner transform for a focused transducer. The spatial and time-dependent part of the Wigner transform in Eq. (61) is given by

$$A_1(r,z,t) = \frac{1}{w^2(z)} \exp\left[-2\alpha^L z + 2z \frac{(2c_L t - z)}{\sigma^2 c_L^2} - 2\frac{r^2}{w^2(z)} \right.$$
$$\left. - 2\left(\frac{t}{\sigma}\right)^2 \right]. \quad (74)$$

Along the transducer axis ($r=0$), this expression becomes

$$A_1(0,z,t) = \frac{1}{w^2(z)} \exp\left[-2\alpha^L z + 2z \frac{(2c_L t + z)}{\sigma^2 c_L^2} - 2\left(\frac{t}{\sigma}\right)^2 \right], \quad (75)$$

which describes the amplitude of the Wigner transform of the mean source field along the z axis. The contour of this amplitude is shown in Fig. 1 for aluminum. The color bar indicates the amplitude value of the contour. The contour has a peak arrival at the focus ($F = 12$ mm) at $t = 1.91$ μs. The amplitude is maximum around the focal region and approaches zero away from it. The contour is a Gaussian curve along the line defined by $2c_L t - z = 0$.
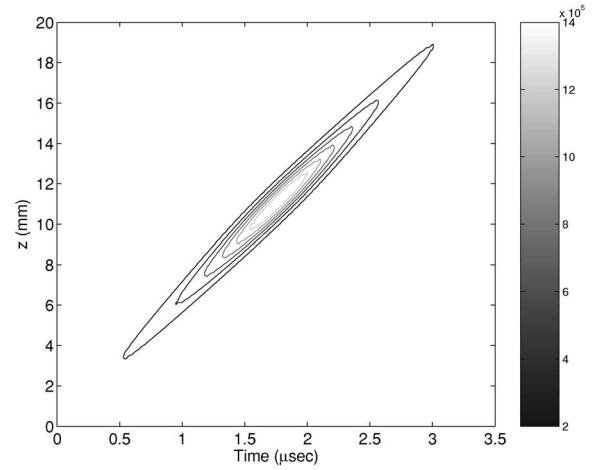
FIG. 1. Contour plot of the $A_1(0,z,t)$, defined by Eq. (75) for aluminum.

Example results for the SSR, Eq. (70) normalized by Φ_0 are shown in Fig. 2. The central frequency of all the input signals is 15 MHz. The SSR given by Eq. (70) depends on the input signal duration σ. The SSR for various input signals of time duration $t = 1$, 2, 3, and 4 μs are shown. From Eq. (70) and the Gaussian beam parameters shown in Eq. (72), the effects of the duration of the input signal on the SSR are shown in Fig. 2. The SSR increases with an increase of the total energy of the input signal. The SSR for 4 μs is almost double that of the 1 μs input signal.

A comparison of the SSR for aluminum and steel is shown in Fig. 3. The amplitude of the singly scattered response increases as the scattering in the material increases. The peak amplitude for steel is higher than for aluminum by more than a factor of 10. The peak arrival time for the two results also differs since the focal depth depends on the sound speed of the respective material.

Finally, the influence of the focal depth is also examined. The SSR for focal depths of 0.012, 0.02, 0.1, and 1 m in an aluminum sample are shown in Fig. 4. As the focal length increases and approaches infinity (similar to a planar transducer), the peak in the SSR shifts and finally vanishes. The convergence of the energy at the focal point results in a
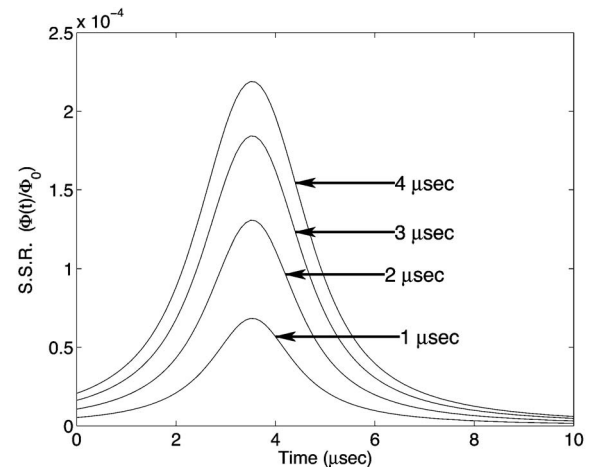
FIG. 2. Normalized singly scattered response (SSR) for input signals of various duration at 15 MHz for aluminum sample.
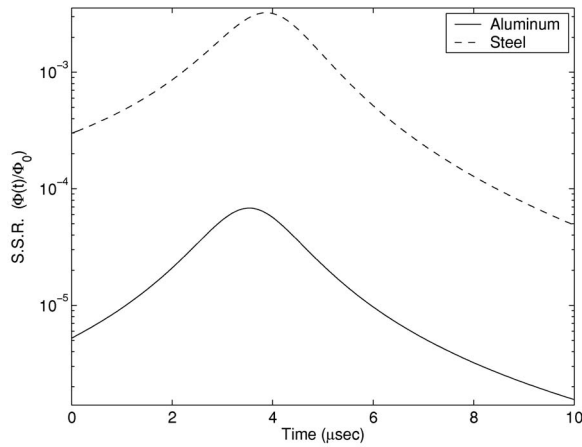
FIG. 3. Normalized SSR in aluminum and steel.

SSR enhancement over that from an unfocused transducer.

IX. SUMMARY

A theoretical model of SSR in polycrystalline materials for longitudinal-to-longitudinal scattering is presented. The mean response of the medium is deduced using the first-order smoothing approximation of the self-energy from the Dyson equation. Then the Green's function covariance is derived from the Bethe-Salpeter equation which includes all orders of scattering information, within the context of the ladder approximation. Then the covariance is simplified within a single scattering assumption in conjunction with the Wigner transform of transducer beam model. The Wigner transform of the beam model is proportional to the energy of the source field in the media as a function of space, time, wave vector, and frequency. Finally, the model derived here is tested for a Gaussian beam model with a paraxial approximation and example results are shown for materials of common interest. The model is tested for various parameters (such as the focal depth and duration of the input signal), to observe their influence on the results.

The model for the SSR derived here can be used for microstructural evaluation in the same way that the Thompson and Gray model[17] has been used: to infer microstructural

information or to identify flaws within a heterogeneous medium. However, one of the impacts of this derivation is associated with the limits of the single-scattering assumption. Since the SSR was derived here within the context of a full multiple scattering formalism, the influence of higher-order scattering can be rigorously examined for realistic measurements (i.e., with a transducer model included in the analysis) as a function of space and time. Thus, experimental data that do not appear to satisfy this assumption may ultimately be understood. Margetan et al.[57] have shown the effects of multiple scattering in backscatter measurements for a strong scattering material. The present model can also predict the response for a pitch-catch scattering measurement setup where the transmitter and the receiver transducer are oriented at an angle relative to one another.[58] The pitch-catch method is mainly used to invoke shear wave propagation in the materials. In addition, the development of new, innovative characterization methods that exploit multiple scattering may be anticipated in the future. Finally, it is expected that the derivation of the Wigner distribution for the focused Gaussian beam developed here may be easily generalized to other transducer models, expanding the range of applicability of this derivation. The wide variety of potential new studies of heterogeneous media is exciting indeed.

FIG. 4. Normalized SSR for various focal depths in aluminum.

[1]A. K. Holmes, R. E. Challis, and D. J. Wedlock, "A wide bandwidth study of ultrasound velocity and attenuation in suspensions: Comparison of theory with experimental measurements," J. Colloid Interface Sci. **156**, 261–268 (1993).

[2]M. S. Greenwood, J. L. Mai, and M. S. Good, "Attenuation measurements of ultrasound in a Kaolin-Water slurry: A linear dependence on frequency," J. Acoust. Soc. Am. **94**, 908–916 (1993).

[3]J. M. Hovem and G. D. Ingram, "Viscous attenuation of sound in saturated sand," J. Acoust. Soc. Am. **66**, 1807–1812 (1979).

[4]C. M. Sayers and R. L. Grenfell, "Ultrasonic propagation through hydrating cements," Ultrasonics **31**, 147–153 (1993).

[5]M. Biot, "Theory of propagation of elastic waves in a fluid saturated porous solid I," J. Acoust. Soc. Am. **28**, 168–178 (1956); **28**, 179–191 (1956).

[6]D. Wu, Z. W. Qian, and D. Shao, "Sound attenuation in a coarse granular medium," J. Sound Vib. **162**, 529–535 (1993).

[7]R. C. Courtney and L. Mayer, "Acoustic properties of fine grained sediments," J. Acoust. Soc. Am. **93**, 3193–3200 (1993).

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

[9]H. Tavossi and F. Cohen-Tenoudji, "Ultrasonic investigation of contact surfaces between grains in random granular media. Effects of variable compression," Proceedings of Ultrasonics International Conference, Vienna, 1993, pp. 347–350.

[10]L. Tsang, J. A. Kong, and T. Habashy, "Multiple scattering of acoustic waves by random distributions of discrete spherical scatterers," J. Acoust. Soc. Am. **71**, 552–558 (1982).

[11]J. Liu, L. Ye, D. Weitz, and P. Sheng, "Novel acoustic excitations in suspensions of hard sphere colloids," Phys. Rev. Lett. **65**, 2602–2605 (1990).

[12]R. L. Weaver, "Diffuse field decay for materials characterization," in *Solid Mechanics Research for Quantitative Non-Destructive Evaluation*, edited by J. D. Achenbach and Y. Rajapakse (Martinus Nijhoff, The Hague,

Ghoshal et al.: Diffuse backscatter in heterogeneous media

1987), pp. 425–434.

[13]C. B. Guo, P. Holler, and K. Goebbels, "Scattering of ultrasonic waves in anisotropic polycrystalline metals," Acustica **59**, 112–120 (1985).

[14]R. L. Weaver, W. Sachse, K. Green, and Y. Zhang, "Diffuse ultrasound in polycrystalline solids," in *Proceedings of Ultrasonics International '91* (Butterworth-Heinemann, Oxford, UK, 1991), pp. 507–510.

[15]R. L. Weaver, "Ultrasonic diffuse field measurements of grain size," in *Non-Destructive Testing and Evaluation in Manufacturing and Construction*, edited by H. L. M. D. Reis (Hemisphere, New York, 1990), pp. 425–434.

[16]R. L. Weaver, "Diffusivity of ultrasound in polycrystals," J. Mech. Phys. Solids **38**, 55–86 (1990).

[17]F. J. Margetan, T. A. Gray, and R. B. Thompson, "A technique for quantitative measuring microstructurally induced ultrasonic noise," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1991), Vol. **10**, pp. 1721–1728.

[18]R. L. Weaver, "Anderson localization of ultrasound," Wave Motion **12**, 129–142 (1990).

[19]R. L. Weaver, "Diffuse waves in finite plates," J. Sound Vib. **94**, 319–335 (1984).

[20]R. L. Weaver, "Laboratory studies of diffuse waves in plates," J. Acoust. Soc. Am. **79**, 919–923 (1986).

[21]J. A. Turner and R. L. Weaver, "Time dependence of multiply scattered diffuse ultrasound in polycrystalline media," J. Acoust. Soc. Am. **97**, 2639–2644 (1995).

[22]J. A. Turner and R. L. Weaver, "Radiative transfer of ultrasound," J. Acoust. Soc. Am. **96**, 3654–3674 (1994).

[23]J. A. Turner and R. L. Weaver, "Radiative transfer and multiple scattering of diffuse ultrasound in polycrystalline media," J. Acoust. Soc. Am. **96**, 3675–3683 (1994).

[24]R. L. Weaver and W. Sachse, "Diffusion of ultrasound in a glass bead slurry," J. Acoust. Soc. Am. **97**, 2094–2102 (1995).

[25]A. Tourin, M. Fink, and A. Derode, "Multiple scattering of sound," Waves Random Media **10**, R31–R60 (2000).

[26]P. Anugonda, J. S. Wiehn, and J. A. Turner, "Diffusion of ultrasound in concrete," Ultrasonics **39**, 429–435 (2001).

[27]A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978).

[28]L. Tsang, J. A. Kong, and R. T. Shin, *Theory of Microwave Remote Sensing* (Wiley-Interscience, New York, 1985).

[29]L. Tsang, J. Kong, and K. Ding, *Scattering of Electromagnetic Waves: Theory and Applications* (Wiley Interscience, New York, 2000).

[30]L. Tsang and J. Kong, *Scattering of Electromagnetic Waves: Advanced Topics* (Wiley Interscience, New York, 2001).

[31]*3D Radiative Transfer in Cloudy Atmosphere*, edited by A. Marshak and A. B. Davis (Springer, Berlin, 2005).

[32]M. I. Mishchenko, L. D. Travis, and A. A. Lacis, *Multiple Scattering of Light by Particles: Radiative Transfer and Coherent Backscattering* (Cambridge University Press, Cambridge, 2006).

[33]B. Davison, *Neutron Transport Theory* (Oxford-Clarendon Press, Oxford, 1957).

[34]K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, 1967).

[35]G. I. Bell and S. Glasstone, *Nuclear Transport Theory* (Van Nostrand, New York, 1970).

[36]L. Ryzhik, G. Papanicolaou, and J. B. Keller, "Transport equations for elastic and other waves in random media," Wave Motion **24**, 327–370 (1996).

[37]N. Trégourès, R. Hennino, C. Lacombe, N. M. Shapiro, L. Margerin, M. Campillo, and B. A. V. Tiggelen, "Multiple scattering of seismic waves," Ultrasonics **40**, 269–274 (2002).

[38]A. E. Malcom, J. A. Scales, and B. A. V. Tiggelen, "Extracting the Green function from diffuse, equipartioned waves," Phys. Rev. E **70**, 015601 (2004).

[39]N. Trégourès and B. A. V. Tiggelen, "Generalized diffusion equation for multiple scattered elastic waves," Waves Random Media **12**, 21–38 (2002).

[40]N. P. Trégourès and B. A. V. Tiggelen, "Quasi-two-dimensional transfer of elastic waves," Phys. Rev. E **66**, 036601 (2002).

[41]S. Chandrasekhar, *Radiative Transfer* (Dover, New York, 1960).

[42]L. Cohen, *Time-Frequency Analysis* (Prentice Hall, Englewood Cliffs, NJ, 1995).

[43]E. Wigner, "On the quantum correction for thermodynamic equilibrium," Phys. Rev. **40**, 749–759 (1932).

[44]R. L. Weaver, "The effect of an undamped finite degree of freedom 'fuzzy' substructure: Numerical solutions and theoretical discussion," J. Acoust. Soc. Am. **100**, 3159–3164 (1996).

[45]U. Frisch, "Wave propagation in random media," in *Probabilistic Methods in Applied Mathematics*, edited by A. T. Barucha-Reid (Academic, New York, 1968), Vol. **1**, pp. 75–198.

[46]P. J. Loughlin and L. Cohen, "A Wigner approximation for wave propagation (L)," J. Acoust. Soc. Am. **118**, 1268–1271 (2005).

[47]L. Borcea, G. Papanicolaou, and C. Tsogka, "Adaptive interferometric imaging in clutter and optimal illumination," Inverse Probl. **22**, 1405–1436 (2006).

[48]D. T. Smithey, M. Beck, and M. G. Raymer, "Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum," Phys. Rev. Lett. **70**, 1244–1247 (1993).

[49]A. Wax, S. Bali, and J. E. Thomas, "Time-resolved phase-space distribution for light backscattered from a disordered medium," Phys. Rev. Lett. **85**, 66–69 (2000).

[50]F. Reil and J. E. Thomas, "Observation of phase conjugation of light arising from enhanced backscattering in a random media," Phys. Rev. Lett. **95**, 143903 (2005).

[51]R. B. Thompson and T. A. Gray, "A model relating ultrasonic scattering measurement through liquid-solid interfaces to unbounded medium scattering amplitudes," J. Acoust. Soc. Am. **74**, 1279–1290 (1983).

[52]R. B. Thompson and E. F. Lopes, "The effects of focusing and refraction on Gaussian ultrasonic beams," J. Nondestruct. Eval. **4**, 107–123 (1984).

[53]B. D. Cook and W. J. Arnoult III, "Gaussian-Laguerre/Hermite formulation for the nearfield of an ultrasonic transducer," J. Acoust. Soc. Am. **59**, 9–11 (1976).

[54]A. D. Kim and A. Ishimaru, "Optical diffusion of focused beam wave pulse in discrete random media," Proc. SPIE **3914**, 423–434 (2000).

[55]J. A. Turner, "Elastic wave propagation and scattering in heterogeneous, anisotropic media: Textured polycrystalline materials," J. Acoust. Soc. Am. **106**, 541–552 (1999).

[56]R. B. Thompson, "Elastic wave propagation in random polycrystals: Fundamentals and application to nondestructive evaluation," in *Imaging of Complex Media with Acoustic and Seismic Waves, Topics in Applied Physics*, edited by M. Fink, W. A. Kuperman, J. Montagner, and A. Tourin (2002), Vol. **84**, pp. 233–257.

[57]F. J. Margetan, P. Haldipur, L. Yu, and R. B. Thompson, "Looking for multiple scattering effects in backscattered ultrasonic grain noise from jetengine nickel alloys," *Review of Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti, (American Institute of Physics, Melville, New York, 2005), Vol. **24**, pp. 75–82.

[58]F. J. Margetan, R. Roberts, and R. B. Thompson, "Estimates of signal-to-microstructural-noise ratios in ultrasonic inspections of metals," *Review of Quantitative Nondestrictive Evaluation*, edited by D. O. Thompson and D. E. Chimenti, (American Institute of Physics, Melville, NY, 2006), Vol. **25**, pp. 1193–1200.

[59]*Matlab*, *The Language of Technical Computing* (The Math Works, Inc., Natick, MA, 2001).

[60]Wolfram Research, I., *Mathematica*, version 4.0 (Wolfram Research, Inc., Champaign, IL, 1999).

Rectangular cymbal arrays for improved ultrasonic transdermal insulin delivery

Joseph Luis
Graduate Program in Acoustics, The Pennsylvania State University, University Park, Pennsylvania 16802

Eun Joo Park
Department of Bioengineering, The Pennsylvania State University, University Park, Pennsylvania 16802

Richard J. Meyer, Jr.
Applied Research Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802

Nadine Barrie Smith[a)]
Graduate Program in Acoustics, The Pennsylvania State University, University Park, Pennsylvania 16802

Circular cymbal ultrasound arrays have been shown to be effective in delivering therapeutic levels of insulin in rats, rabbits, and pigs. To improve delivery efficiency, a rectangular cymbal design was desired in order to achieve a broader spatial intensity field without increasing the size of the device or the spatial-peak temporal-peak intensity (I_{SPTP}). With a similar intensity (50 mW/cm^2), the goal was to determine if the 3×1 rectangular cymbal array could perform significantly better than the 3×3 circular array for glucose reduction in hyperglycemic rabbits. Rabbit experiments were performed using three groups: nonsonicated control ($n=3$), ultrasound exposure using a circular cymbal array ($n=3$), and ultrasound exposure using a rectangular cymbal array ($n=3$). Rabbits were anesthetized and a water tight reservoir that held the insulin was fastened on the rabbit's thigh. At the beginning of the experiment and every 15 min for 90 min, the blood glucose level was determined. For comparison between individual rabbits, the absolute level is normalized by subtracting out the baseline in order to arrive at the change in glucose level. For the control group, the normalized glucose level increased (more hyperglycemic) to $+80.0 \pm 28.8$ mg/dl (mean$\pm$SEM). Using the circular array, the glucose level decreased to -146.7 ± 17.8 mg/dl at 90 min. However, using the rectangular cymbal array, the glucose decreased faster and to a level of -200.8 ± 5.9 mg/dl after 90 min. These results indicated the feasibility of the rectangular cymbal array as an improved device for drug delivery. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2769980]

I. INTRODUCTION

Noninvasive methods of drug delivery have been studied for some time to eliminate the pain induced by methods such as injections.[1–5] In particular, a large number of type I and type II diabetics require blood glucose level regulation by injections. Studies have shown that ultrasound mediated transdermal drug delivery offers promising potential for noninvasive drug administration.[6–9] Ultrasound accomplishes this by making the skin permeable, which allows drugs to more readily diffuse. The working principle, although not completely understood, has been suggested to be the result of cavitation.[10–13] Low frequency ultrasound in the vicinity of 20 kHz is capable of generating microbubbles in the water and tissue layers. These bubbles allow water channels to be produced within the cellular lipid bilayers. The resulting disorder created in the *stratum corneum* facilitates the crossing of a hydrophilic drug or molecule. To this end, the number of drugs and compounds that have been shown to transdermally cross skin via ultrasound is ever increasing.[9] This work further explores the use of ultrasonic treatments as a promising method for delivering therapeutic levels of drugs to patients.

In addition to having a system to sonicate and deliver drugs transdermally, it would be useful to have a portable delivery device.[7] A transducer array has been developed that can produce the necessary intensity levels to increase skin permeability and maintain a low profile.[14] The low-profile "ultrasound patch" uses the cymbal, which is a flextensional transducer made of a circular piezoelectric ceramic sandwiched between two metal end caps.[14–17] The small radial displacement of the ceramic is mechanically transformed into larger displacements by the metal caps. Cymbal transducers can be arranged into arrays of any size and maintain a low profile, be lightweight, and be conformable.

Various animal tests using rats, rabbits, and pigs have been performed to demonstrate the feasibility of the 2×2 and 3×3 circular cymbal arrays in transdermal insulin delivery systems.[18–20] These circular cymbal array designs were able to produce significant reductions in glucose levels. Specifically, in previous rat experiments, the 2×2 array was used to deliver 20 kHz ultrasound operating at a spatial-peak

---
[a)]Author to whom correspondence should be addressed at College of Engineering, Department of Bioengineering, The Pennsylvania State University, 218 Hallowell Building, University Park, PA 16802. Electronic mail: nbs@engr.psu.edu

temporal-peak intensity (I_{SPTP}) of 100 mW/cm^2 for 90 min. Xylazine was used as a general anesthetic as well as to cause temporary hyperglycemia.[21,22] Results indicated a blood glucose decrease (mean±standard deviation) of 297.7±52.8 mg/dl after 90 min.[18] Since therapeutic levels of insulin would need to be delivered to larger animals, a 3 ×3 circular array configuration was built. Experiments on rabbits were performed under the same intensity and pulse conditions using the slightly larger 3×3 array. For rabbits, the blood glucose decrease was 136.1±26.1 mg/dl after 90 min.[19]

Although the cymbal transducer allows for a thin device, its circular geometry is not optimal for array packaging. For a portable insulin delivery device, we would also prefer a cymbal transducer that is electrically more efficient at the operating frequency in order to minimize battery drain. One improvement can be made by bringing the ceramic and cap resonances closer together. Doing so increases the displacement of the caps and increases the electrical efficiency.

In an ultrasound device for transdermal drug delivery, a larger spatial average intensity can offer higher drug diffusion rates with a larger permeabilized skin area. Previous 2 ×2 and 3×3 cymbal arrays have shown spatial intensity distributions with high intensities at the center of the device and a rapid intensity decrease away from the spatial peak.[19,23] This type of spatial intensity distribution results in larger arrays that are not able to adequately improve the insulin delivery rate. By exciting higher order modes in the cymbal cap, we can broaden the spatial extent of the intensity field. To achieve a greater electrical efficiency and obtain a broader spatial intensity field at a 1 mm distance from the transducer, a rectangular cymbal was designed and evaluated.[24] The rectangular array was no larger than the 3 ×3 circular array previously tested and *in vivo* rabbit experiments were performed herein to verify the effectiveness of the device.

II. MATERIAL AND METHODS

A. Transducer development

Details regarding the development of the cymbal transducer have been described elsewhere.[14,24,25] The cymbal transducer is a novel flextensional transducer capable of producing very low frequencies. The cymbal transducer has a compact, lightweight structure with an adjustable resonance frequency. In the cymbal transducer design, the caps on the lead zirconate-titanate (PZT) ceramic contained a shallow cavity beneath its inner surface. The fundamental mode of vibration is the flexing of the end caps caused by the radial motion of the ceramic. Therefore, the overall displacement of the device is a combination of the axial motion of the disk plus the radial motion amplified by the end caps.

The circular cymbal arrays use the first resonance frequency of the caps for acoustic radiation.[15,16,25–27] In order to be near a low frequency, it was necessary to increase the cap size such that the close operation frequency excites higher order modes in the cap. Higher order modes will create a more desirable spatial intensity profile. Yet, increasing the

overall size of a circular ceramic decreases the first ceramic resonance frequency. Radial modes are governed by

$$J_1(\eta) = 1 - \sigma^p, \tag{1}$$

where J_1 is the Bessel function of the first kind of first order, η is an argument proportional to the frequency, and σ^p is a planar Poisson's ratio. Here σ^p is defined by $-s_{12}^E/s_{11}^E$, where s^E is the compliance under zero electric field.[28–30] The first root of Eq. (1) places the ceramic resonance of the circular cymbal near 100 kHz, which is far from the 20 kHz first resonance of the cap. A larger cap with a rectangular ceramic geometry resolves this issue. The rectangular cymbal operates in the 3-1 ceramic mode[31] where the longitudinal strain is given by

$$S_s(x) = d_{31}E_0\left[\cos(kx) + \tan\left(\frac{kl}{2}\right)\sin(kx)\right], \tag{2}$$

where d is the piezoelectric strain constant (C/N), E_0 is the electric field applied (V), $k=\omega/c$, c is the speed of sound in the long bar (m/s), and l is the length of the bar (m). The expression for the admittance of the 3-1 bar is shown as

$$Y = \frac{j\omega w d_{31}^2}{s_{11}^E kt}\sin(kl)\left[1 + \tan^2\left(\frac{kl}{2}\right)\right] + \frac{j\omega wl}{t}\left[\varepsilon_{33}^T - \frac{d_{31}^2}{s_{11}^E}\right], \tag{3}$$

where w is the width of the bar (m) and t is the bar thickness (m), and ε^P is the permittivity under zero stress. From Eqs. (2) and (3) the ceramic resonances now occur when $kl=(2n-1)\pi$, giving a first ceramic resonance of $f_r=c/2l$. A bar length of 50 mm, similar to 3×3 circular array dimensions, results in a first resonance of 32 kHz. The system can now be operated at the ceramic resonance instead of the cap resonance giving maximum longitudinal strain of the bar. It was desired to design the rectangular array to have the same area as the 3×3 circular array. In order to have a rectangular array of the same size, a 3×1 array was made with element dimensions of $50\times13\times3$ mm^3.

B. Transducer modeling

A finite element model of a single element rectangular transducer was developed to predict the cap displacement and electrical impedance. The finite element analysis code ATILA (ISEN, Lille, France, Hamonic) was used. ATILA is a particularly useful tool for piezoelectric and magnetostrictive systems[32] and has been previously used for the simulation of these flextensional transducers. Figure 1(a) shows a 1/8th model of a single rectangular transducer element and the predicted cap displacement when operated at the first ceramic resonance. Modeled at 32 kHz, Fig. 1(b) is a contour plot showing the axial displacement of the rectangular cymbal caps at the first resonance frequency. The color bar represents the displacement in meters.

C. Rectangular array construction

Construction of the rectangular cymbal transducers was done by hand. It is important to maintain symmetry when fabricating the transducers. Rectangular Navy Type I 402

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Luis *et al.*: Rectangular arrays for insulin delivery   2023
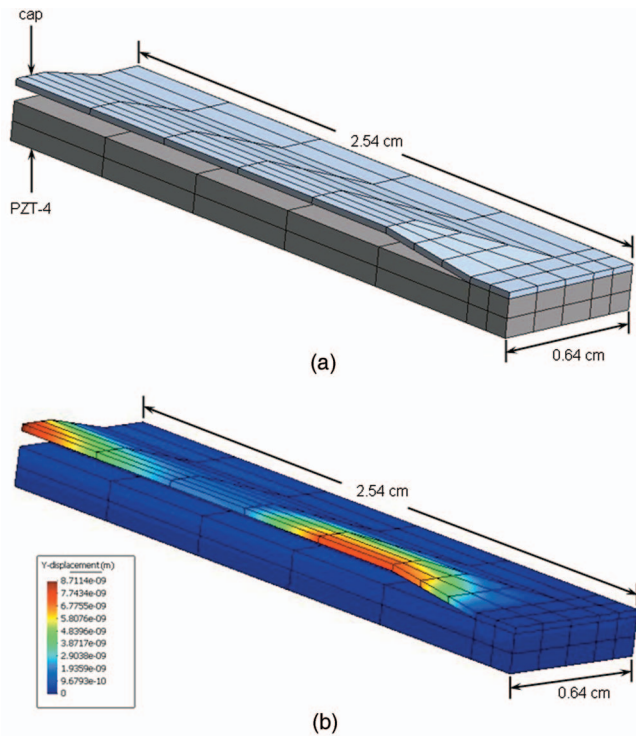
FIG. 1. (a) ATILA finite element analysis software was used to establish a 1/8th model of a single element rectangular cymbal transducer. (b) Using ATILA, the predicted cap displacement when operated at the first ceramic resonance was modeled. The axial displacement of the rectangular cymbal caps is indicated with the color bar and the displacement is in meters.

ceramics (Piezokinetics, Inc., Bellefonte, PA) were poled in the thickness direction to use in the 3-1 bar configuration. The titanium caps were punched from 0.25-mm-thick sheets (Item 10385, Alfa Aesar, Ward Hill, MA). Initially, titanium was cut into strips the same width and length as the rectangular ceramic. The cap was shaped by pressing the cap using a specially machined shaping die. One cap at a time was placed in the shaping die and pressed (Model 3012, Carver, Inc., Wabash, IN) at 60 kpsi. Cap heights were measured at the center of the cap using a precise dial gauge (Model 254MZ-300, L. S. Starrett Co., Athol, MA); for symmetry, caps with the closest cap heights were paired together. Caps were bonded to the ceramics using Eccobond® epoxy (Emerson & Cuming, Billerica, MA) base and 15 LV catalyst in a 3:1 ratio. After a rectangular cymbal transducer was constructed [Fig. 2(a)], the admittance (magnitude and phase) was measured in air using a network analyzer (Agilent E5100A, Santa Clara, CA) for comparison to the ATILA model as a function of frequency.

Multielement arrays were made by electrically connecting the rectangular transducers and potting the elements in polyurethane (URALITE® polymer, FH 3550, H.B. Fuller, St. Paul, MN). The two part polyurethane with a 90A shore hardness was mixed, degassed, and poured on the array, which was set in a mold. To compare the delivery efficiency of the 3×3 circular design to the new 3×1 rectangular design, a previously constructed circular array was used.[19] Figure 2(b) shows a 3×3 array (left) with nine circular cymbal transducers compared to a 3×1 array using three rectangular cymbal transducers (right).

FIG. 2. (Color online) (a) A single element rectangular cymbal transducer was constructed from titanium cap molded from a shaping die. Caps were bonded to the ceramics using epoxy. (b) Multielement arrays were made by electrically connecting the transducers and potting them in polyurethane. The 3×1 rectangular array (right) was designed to be comparable in size to the 3×3 circular array (left).

To drive the rectangular and cymbal arrays, a (rf) signal was generated by a frequency pulse/function generator (Agilent 32250A, Palo Alto, CA) and amplified by a rf amplifier (Model 40A12, Amplifier Research, Souderton, PA). The electrical impedance of both the rectangular and circular array was tuned to the output impedance of the amplifier using an external inductive-capacitive (LC) π network. Pulse period, duty cycle, and on-time of the rf signal from the frequency generator was monitored using an oscilloscope (Tektronix 2213A, Beaverton, OR). For the experiments, the signal generator operated at approximately 32 kHz with pulse durations of 200 ms and pulse repetition period of 1 s (i.e., 20% duty cycle); the amplifier gain was set to 50 dB. Pulsed ultrasound was used to avoid damaging heat build-up to either the array's piezoceramic or the animal's skin.

D. Far field measurements

As a method of comparing the transducer efficiency, far field measurements in water were acquired at the anechoic

Luis *et al.*: Rectangular arrays for insulin delivery

acoustic test facilities in the Applied Research Laboratory (State College, PA). Transmitting voltage response (TVR), complex impedance, and directivity patterns were measured. Transmit system performance in terms of volt-amperes per acoustic watt was calculated. TVR is a ratio of the sound pressure at a distance of 1 m in a specified direction to the voltage applied. The reference voltage is 1 V and the ratio is expressed in dB re 1 μPa m/V. In an ideal case, all the electrical power is converted to acoustic power. However, in this particular system, the transducer behaves like a capacitor and the conversion of electrical to acoustic power is not ideal. The volt-ampere per acoustic watt ratio (V A/P_{acs}) accounts for both real and reactive electrical power and a ratio of one would imply 100% efficiency. For the cymbal transducers used in previous arrays, a typical untuned value of V A/P_{acs} is on the order of 100.

E. Near field exposimetry

Of interest is the intensity profile close to the arrays for potential bioeffects. For possible practical clinical use, the transducer face will be approximately 1 mm away from the skin. At that distance it is necessary to know how the acoustic intensity field is distributed spatially. The intensity was determined according to exposimetry guidelines previously established.[33,34] For the acoustic field at a plane 1 mm from the transducer face, the ultrasonic intensities from the arrays were determined with a calibrated miniature (4 mm diameter) omnidirectional reference hydrophone (Model TC4013, S/N: 5199093, RESON, Inc., Goleta, CA). The arrays were submerged in a water tank ($51 \times 54 \times 122$ cm^3), which was made almost anechoic by placing 1.27-cm-thick rubber sound absorbing material around its wall. To minimize cavitation effects, a custom made degasser reduced the dissolved oxygen content of the distilled water to 1 to 2 ppm. Pulse period, duty cycle, and on-time of the signal from the frequency generator and hydrophone were acquired using an Agilent 54622A 100 MHz digitizing oscilloscope (Agilent, Palo Alto, CA).

Precise, computer controlled positioning of the hydrophone was performed by a Velmex Positioning System (Velmex Inc., East Bloomfield, NY). Pressure waves detected by the hydrophone were recorded by a digitizing oscilloscope. A computer-controlled exposimetry positioning system was used for automated scanning. The scanning step size for each device was 1 mm and the scanning area was 90×90 mm^2. Spatial peak-temporal peak intensities (I_{sptp}) were determined over a plane 1 mm from the array face using the hydrophone based on three to five scannings of the array for a mean and standard error of the intensity results. Exposimetry field maps were normalized and plotted in decibels with respect to the highest intensity (I_{sptp}) in order to show the spatial distribution of intensity from the maximum radiation point.

F. Animal experiments

Ultrasonic assisted transdermal insulin delivery tests on animals were performed to determine the effectiveness of the two exposure array designs. Experimental protocol for the

(a)

(b)
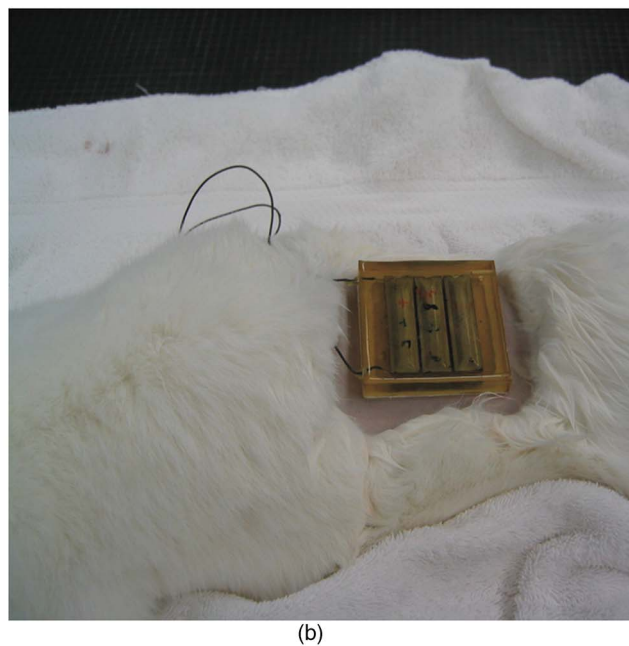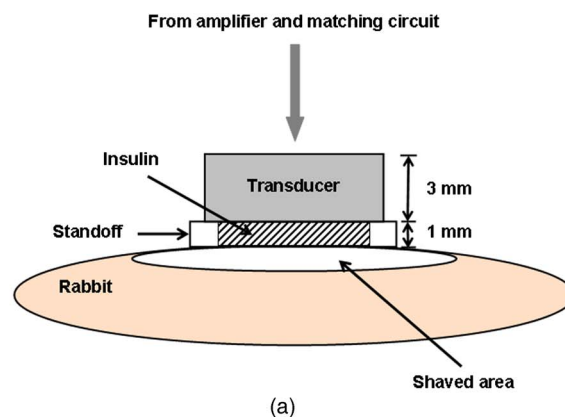
FIG. 3. (Color online) (a) Illustration and (b) photograph of the animal setup for the transdermal insulin transport experiments. With the rabbit placed in a lateral recumbent position, a 1-mm-thick, water-tight standoff filled with insulin was arranged between the thigh and the array.

use of rabbits was reviewed and approved by the Institutional Animal Care and Use Committee at the Pennsylvania State University. New Zealand White rabbits (3.2–4.7 kg) were divided into the three groups (control, 3×3 circular array, and 3×1 rectangular array) and nine experiments were performed with $n=3$ animals in each group. Each animal was anesthetized with a combination of ketamine hydrochloride (40 mg/kg intramuscularly, Ketaject®, Phoenix, St. Joseph, MO) and sodium xylazine (10 mg/kg intramuscularly, Xyla-Ject®, Phoenix, St. Joseph, MO). The hair of the thigh area was shaved; a depilatory agent was very briefly applied to the skin to remove any remaining hair and thoroughly cleaned.

In addition to its use as a general anesthetic, xylazine was used to cause temporary but sustained hyperglycemia (up to 12 h) in rabbits.[21,22] With a rabbit placed in the lateral recumbent position [Fig. 3(a) diagram and Fig. 3(b) photo], a 1-mm-thick, water-tight, standoff was fixed between the thigh and the array (both circular and rectangular design) using double-sided tape (3M, St. Paul, MN). Through small

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Luis et al.: Rectangular arrays for insulin delivery   2025

holes in the back of the arrays, the reservoir inside of the standoff was filled with 4 ml of insulin (Humulin® R, rDNA U-100, Eli Lilly Co., Indianapolis, IN). The insulin was diluted with a saline solution to prepare 50 units/ml of insulin. Previous experiments have shown that this dilution does not affect the overall results.[35] Care was taken to remove all bubbles from the solution in the reservoir between the thigh and the arrays.

Prior to beginning the experiment, 0.3 ml of blood was collected from the ear vein of each rabbit for a baseline blood glucose analysis. The glucose level (mg/dl) in the rabbit blood was determined using an ACCU-CHEK™ Instant® (Roche Diagnostics Co., Indianapolis, IN) blood glucose monitoring system. Multiple samples (three to six each time) were taken from each rabbit for the baseline and at every subsequent 15 min interval for a total of 90 min. The time from the initial injection of the ketamine-xylazine until the baseline glucose measurement was no greater than 20 min. For comparison between individual rabbits, the absolute level is normalized by subtracting out the baseline in order to arrive at the change in glucose level.

For each rabbit, the entire experiment lasted a total of 90 min. The control group ($n=3$) used insulin in the reservoir without any ultrasound operating (designated the "control—insulin without ultrasound"). For the exposure groups, the second group ($n=3$) used insulin in the standoff with ultrasound exposure using the 3×3 circular array ($I_{sptp}=50.1 \pm 1.3$ mW/cm^2), whereas the last group ($n=3$) used insulin with ultrasound exposure using the 3×1 rectangular array operated at a comparable intensity ($I_{sptp}=50.8 \pm 0.6$ mW/cm^2). For all three groups, the standoff reservoir with the insulin was removed at 60 min, although glucose measurements continued until 90 min from the start. After the experiments, the rabbits were carefully observed until recovery from the anesthesia.

Statistical analysis was performed using Microsoft EX-CEL® (Microsoft Corp., Redmond, WA). The blood glucose versus time data for each group was analyzed as mean and standard error ($x \pm$ SEM). An analysis of variance (ANOVA) was used to analyze the statistical significance of the differences among the means of groups. The p value was used to determine if the between-group differences are significantly greater than those attributable to chance. For the data, a single was used if the p value was less than the 0.05 level of significance.

III. RESULTS

A. Exposimetry

After bonding the cap to the piezoceramic, each rectangular transducer's admittance in air was measured and compared to the ATILA predicted admittance. Figure 4(a) shows a typical measured and calculated admittance (magnitude and phase) for a single element rectangular transducer in air prior to potting in polyurethane. TVR and a volt-ampere per acoustic watt ratio (V A/P_{acs}) were determined for the 3 $\times$ 1 rectangular arrays and the 3×3 circular arrays. Figure 4(b) shows the measured TVR for the 3×3 circular and the
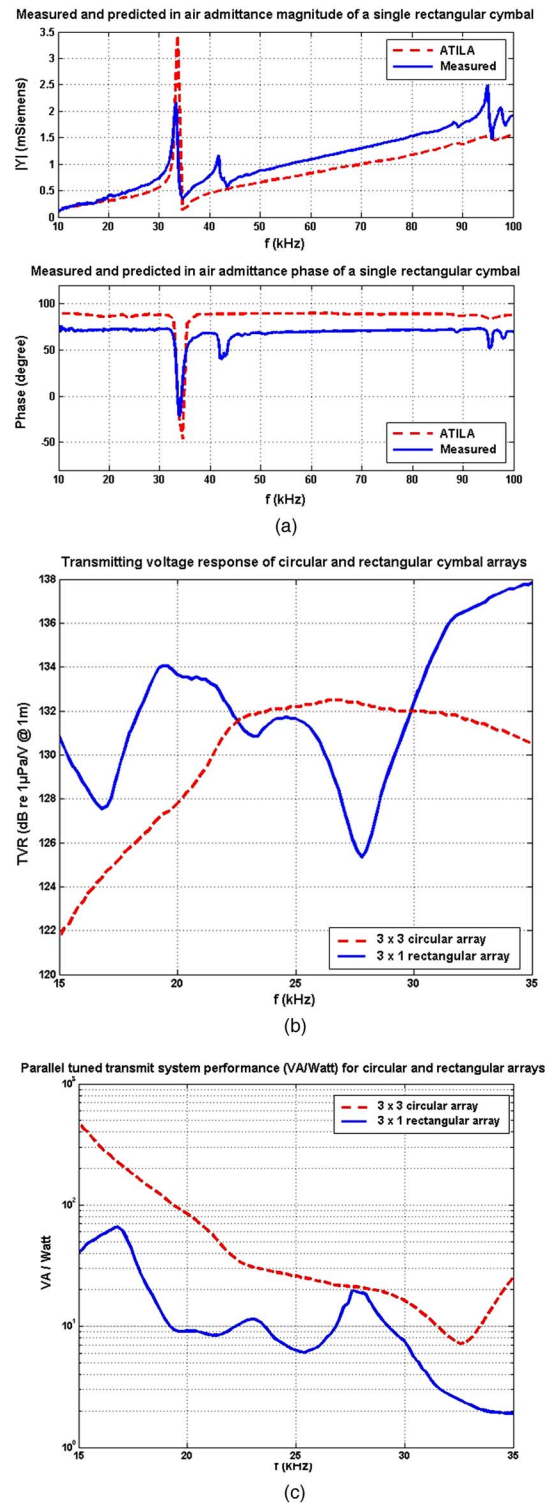
FIG. 4. (Color online) (a) Measured and ATILA predicted admittance (magnitude and phase) of a rectangular cymbal transducer in air. (b) Transmitting voltage response (TVR) comparison of a 3×1 rectangular and 3×3 circular array. Except for the dips, the rectangular array has a greater TVR than the circular array. The dips in the TVR for the rectangular array were due to asymmetrical modes in the device possibly caused by uneven bonding. (c) Parallel tuned V A/P_{acs} ratios for both arrays, which indicates an improved efficiency for the rectangular array over the circular array.

3×1 rectangular arrays with dips near 17, 23, and 29 kHz in the TVR for the rectangular array, which are due to asymmetrical modes in the device.

Over the frequency range of interest including the resonance of the arrays (circular, 20 kHz; rectangular, 32 kHz), the V A/P_{acs} ratio was compared between the 3×3 circular and the 3×1 rectangular arrays. For single frequency operation the arrays can be tuned to obtain a high efficiency at the operating frequency. For this particular work, the arrays were parallel tuned. An inductor was placed in parallel with the device and was designed to resonate at a user-specified frequency with the known clamped capacitance of the array. Figure 4(c) shows the measured tuned V A/P_{acs} ratios for the 3×3 circular and the 3×1 rectangular arrays. Manifested as dips in the TVR in Fig. 4(b), there are peaks in Fig. 4(c) that correspond to asymmetrical modes in the transducer due to small asymmetries in the device introduced during the construction process. Moreover the V A/P_{acs} ratios for the 3×3 circular is lower than the ratio for 3×1 rectangular arrays indicating an improved efficiency for the rectangular array compared to the circular array.

To compare drug delivery efficiency of the rectangular and the circular designs, both arrays were driven such that the I_{sptp} would be similar. From multiple exposimetry scans, the intensity of the 3×1 rectangular array was $I_{\text{sptp}} = 50.8 \pm 0.6$ mW/cm^2 while the intensity of the 3×3 circular array was $I_{\text{sptp}} = 50.1 \pm 1.3$ mW/cm^2. Exposimetry field maps were plotted in decibels with respect to the highest intensity (I_{sptp}) in order to show spatial distribution of the intensity field from the maximum radiation point. Using similar I_{sptp} driving conditions, the intensity field was determined in a plane 1 mm from the array face for both the rectangular and the circular designs. Figure 5(a) shows a typical two-dimensional scanning plot of the temporal peak intensity for the 3×3 circular cymbal transducer over a 90×90 mm^2 area with 1 mm steps. The grayscale bar on the side of the graph illustrates the normalized temporal peak intensity change in decibels. Under similar driving conditions, a 90×90 mm^2 two-dimensional plot of the temporal peak intensity was acquired for the rectangular array [Fig. 5(b)]. As seen in Fig. 5, even though two arrays have similar I_{sptp} values, the intensity distributions were different with the 3×1 rectangular array having a larger area of high intensity.

B. Animal experiment results

Results of the ultrasonic transdermal insulin delivery in rabbits for the three groups are graphed (Fig. 6) as the change in the blood glucose level during the 90 min experiment. After the rabbits were anesthetized, the average initial glucose level of the nine rabbits was 241.0 ± 1.9 mg/dl. Generally for normal rabbits, the blood glucose level is approximately $100-135$ mg/dl.[21,22] For comparison between the rabbits, the change in the blood glucose level was normalized to a baseline, which was the initial glucose level for each rabbit.

Data were graphed as the mean$\pm$standard error ($x \pm$SEM) of each group. For the control group (Fig. 6, diamonds, "insulin without ultrasound"), the glucose level increased to 63.7 ± 22.8 mg/dl after 30 min compared to the initial baseline. Over the 90 min experiment the slope of this blood glucose increase was $+45.7$ mg/dl/h ($r^2 = 0.7$) to a
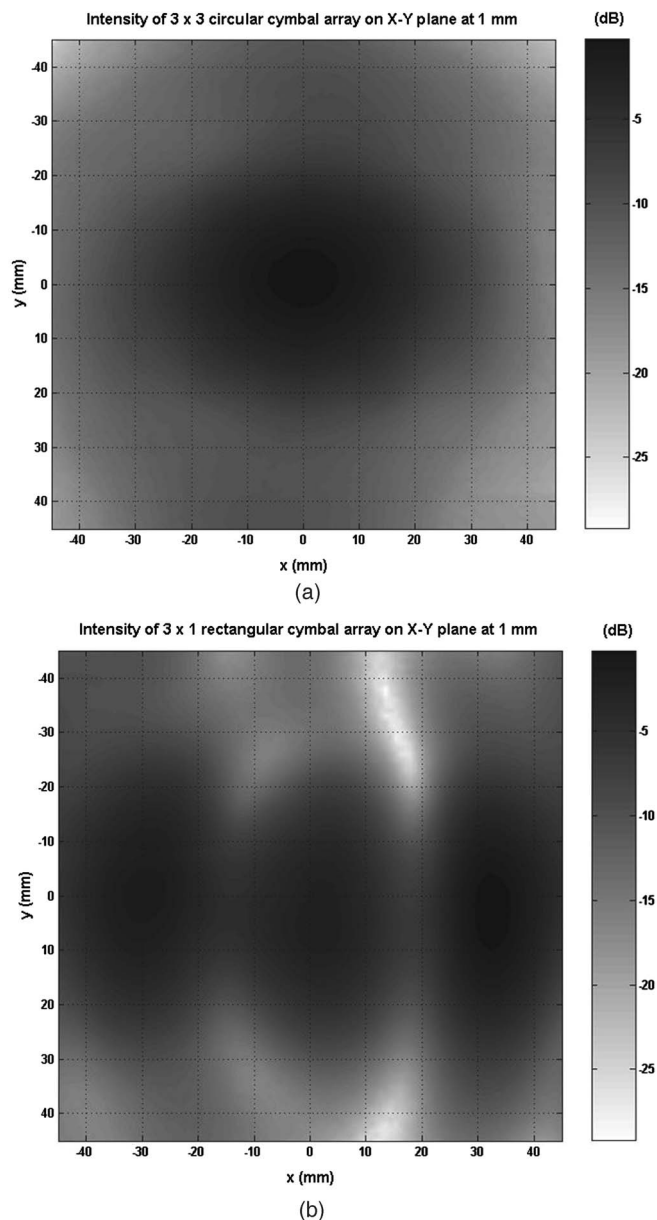
FIG. 5. Temporal peak intensity profiles of a (a) 3×3 circular and (b) 3×1 rectangular array at a plane 1 mm from the array face. With a similar spatial peak temporal peak intensity for both arrays (≈ 50 mW/cm^2), it is evident that the rectangular array produces a broader spatial intensity field pattern than the circular array. The grayscale bar in decibels indicates the intensity field normalized to the peak intensity.

maximum of $+80.0 \pm 28.8$ mg/dl. For the ultrasound and insulin exposure group using the 3×3 circular array (Fig. 6, squares), the glucose level decreased to -26.7 ± 20.7 mg/dl in 30 min and -100.8 ± 25.1 mg/dl in 60 min from the initial baseline level. After the 3×3 circular array and the insulin reservoir were removed at 60 min, the glucose level continued to decrease to -146.7 ± 17.8 mg/dl at 90 min. Use of the 3×1 rectangular array shows a shows a similar trend but larger magnitude in the decrease in blood glucose levels in the rabbits (Fig. 6, triangles). In comparison to the circular array results at 30 min, use of the 3×1 rectangular array showed a decrease in the blood glucose to -64.3 ± 15.1 mg/dl and to -157.2 ± 9.7 mg/dl at 60 min. For
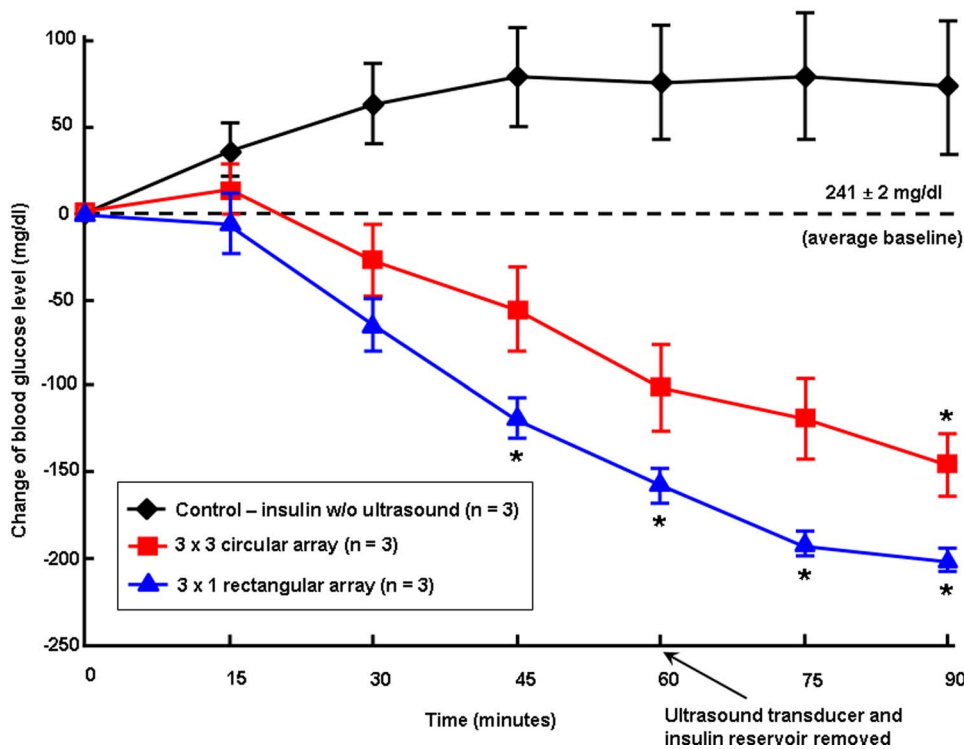
FIG. 6. (Color online) Over a period of 90 min for the control group, the average glucose level increased (i.e., more hyperglycemic) to a maximum of +80.0 mg/dl. While the blood glucose level of the rabbits decreased from the insulin with ultrasound exposure ($I_{sptp}=50$ mW/cm^2) using both arrays, the decrease was greater using the 3×1 rectangular array. Statistically significant differences (p values <0.5, indicated by asterisks) were observed using the rectangular array at 45, 60, 75, and 90 min, while the only statistically significant difference observed using the circular array occurred at 90 min.

the final reading at 90 min the blood glucose level was -200.8 ± 5.9 mg/dl using the rectangular array.

To determine the statistical significance between the control and ultrasound exposure groups using either the circular or rectangular array results in Fig. 6 an ANOVA was used to analyze these data. Asterisks adjacent to a blood glucose-time data point were used if there existed statistically significant difference between the control and exposure data. Comparison of the control with the 3×3 circular array data indicated that there was a statistically significant difference between the groups at 90 min ($p=0.04^*$). However, comparison between control and the 3×1 rectangular array groups indicated that a statistically significant difference existed between these data at 45, 60, 75, and 90 min with p values of 0.02^*, 0.02^*, 0.01^*, and 0.01^*, respectively.

An examination of the rabbit's skin was performed after ultrasound exposure to check for discoloration or visible lesions on the skin surface. The skin did not indicate any noticeable damage or change immediately after the stand-off was removed and for a period of two weeks after the experiments.

IV. DISCUSSION

Cymbal designs have been shown to be effective at delivering therapeutic levels of insulin based on the decrease in blood glucose level after ultrasound exposure *in vivo*. A new cymbal transducer design with a rectangular shape was modeled, constructed, and tested as an improvement over the circular design currently under investigation.

Several methods were used to initially evaluate the design. Using the ATILA finite element analysis software, the in-air admittance agreed well between the model and the measured magnitude and phase result [Fig. 4(a)]. Individual transducers were connected in an array pattern and encased

in polyurethane for in-water electrical efficiency evaluation. For potential future application as a portable system, an untuned device would drain a battery quickly. At single frequency operation, the arrays can be tuned to obtain a high efficiency at the operating frequency. For this particular work, both arrays were parallel tuned. An inductor was placed in parallel with the device and was designed to resonate at a specified frequency with the known clamped capacitance of the array. By doing so, the efficiency was increased in the operating band for both arrays [Fig. 4(c)] compared to an untuned array (graph not shown). Additionally, the TVR for the 3×3 circular array was compared to that of the 3×1 rectangular array [Fig. 4(b)]. From Fig. 4, there are dips in the TVR for the rectangular array that are due to asymmetrical modes in the device. These are due to asymmetries in the device caused by assembling the transducers by hand. Over the rest of the band the rectangular arrays outperform the previous circular cymbals. Yet moreover, the rectangular array generally appears to be more efficient than the circular array except where the V A/P_{acs} plot has peaks due to the asymmetries. For this delivery system, the arrays are intended to operate at a single frequency and the asymmetrical modes are of minimal consequence. In order to use a large frequency band, the manufacturing process needs to be improved to eliminate asymmetries that occur.

Beyond electrically evaluating the arrays for potential use as a drug delivery device, the arrays needed to be characterized in terms of the spatial and temporal intensity. For comparison, both arrays were driven such that the I_{sptp} would be similar (≈ 50 mW/cm^2). Although the I_{sptp} values were similar, the spatial distribution of the temporal peak intensity for the rectangular array [Fig. 5(b)] was much larger than the circular [Fig. 5(a)] array. For the 3×3 circular array, the nine individual transducers appear to be operating as a large

plane-piston transducer with a new effective aperture instead of nine circular transducers. In contrast, the rectangular array shows three peaks giving rise to an intensity field more widely distributed. This larger field can be used to permeabilize a larger area for increased drug delivery.

To evaluate this speculation, experiments were performed using hyperglycemic rabbits. Comparable to previous experiments using the cymbal array and rabbits,[19,23] the 3 × 3 circular array decreased the blood glucose level to a statistically significant level (-146.7^* mg/dl at 90 min) compared to the control group (Fig. 6). However, using the 3 × 1 rectangular array, the blood glucose level decreased faster and to a lower level than the circular design (-200.8^* mg/dl at 90 min). Based on these results, there appears to be a relationship between the spatial extent of the intensity field and the efficiency of insulin delivery. While increasing the spatial-peak intensity may delivery more insulin into the body, this approach also has potential hazardous bioeffects.[36] As with diagnostic ultrasound, the bioeffects and safety of each device needs to be carefully monitored since it will not matter how much of any drug can be transported if the skin is burned, damaged, or the procedure is painful. Therefore increasing the spatial extent of the intensity field may be a better approach than increasing the peak intensity.

In terms of human diabetes, a person is considered diabetic if their blood sugar level is above 140 mg/dl after 8 h of fasting. People without diabetes have fasting sugar levels that generally run between 70 and 110 mg/dl. A glucose of 110–126 mg/dl is classified as impaired fasting glucose, 140–200 mg/dl is impaired glucose tolerance, and greater than 200 mg/dl is considered diabetic.[37–39] In this last situation a diabetic person would need to inject enough insulin to reduce their blood glucose by about 100 mg/dl. Intensities used herein ($I_{sptp} \approx 50$ mW/cm^2) are about one-half of what we have used in previous studies.[19,23,35,40] Compared to our previous circular array results, the rectangular design produced a similar blood glucose decrease but with less peak intensity.

Use of transdermal drug delivery techniques has practical clinical application to medications that need to be injected multiple times either daily or weekly. A recent comprehensive review on ultrasound drug delivery states that "small-sized low-frequency transducers need to be developed so that patients can wear them."[7] We feel that the cymbal meets this need, and the design has a significant amount of trade space for optimization. As with diagnostic ultrasound imaging, drug delivery using therapeutic ultrasound requires a delicate balance between safety and efficacy and requires careful scientific study. The bioeffects and safety of each device need to be carefully monitored since it will not matter how much of any drug can be transported if the skin is burned, damaged, or the procedure is painful.

ACKNOWLEDGMENT

[1]M. R. Prausnitz, "Reversible skin permeabilization for transdermal delivery of macromolecules," Crit. Rev. Ther. Drug Carrier Syst. **14**, 455–483 (1997).

[2]M. R. Prausnitz, "A practical assessment of transdermal drug delivery by skin electroporation," Adv. Drug Delivery Rev. **35**, 61–76 (1999).

[3]F. Montorsi et al., "Transdermal electromotive multi-drug administration for Peyronie's disease: Preliminary results," J. Androl **21**, 85–90 (2000).

[4]Y. Wang, R. Thakur, Q. Fan, and B. Michniak, "Transdermal iontophoresis: Combination strategies to improve transdermal iontophoretic drug delivery," Eur. J. Pharm. Biopharm. **60**, 179–191 (2005).

[5]A. Nanda, S. Nanda, and N. M. Ghilzai, "Current developments using emerging transdermal technologies in physical enhancement methods," Curr. Drug Deliv. **3**, 233–242 (2006).

[6]S. Mitragotri and J. Kost, "Low-frequency sonophoresis: A review," Adv. Drug Delivery Rev. **56**, 589–601 (2004).

[7]W. G. Pitt, G. A. Husseini, and B. J. Staples, "Ultrasonic drug delivery-a general review," Expert. Opin. Drug Deliv. **1**, 37–56 (2004).

[8]M. B. Brown, G. P. Martin, S. A. Jones, and F. K. Akomeah, "Dermal and transdermal drug delivery systems: Current and future prospects," Drug Deliv. **13**, 175–187 (2006).

[9]N. B. Smith, "Perspectives on transdermal ultrasound mediated drug delivery," International Journal of Nanomedicine **2**(4), 1–10 (2007).

[10]S. Mitragotri, D. A. Edwards, D. Blankschtein, and R. Langer, "A mechanistic study of ultrasonically-enhanced transdermal drug delivery," J. Pharm. Sci. **84**, 697–706 (1995).

[11]S. Mitragotri, D. Blankschtein, and R. Langer, "An explanation for the variation of the sonophoretic transdermal transport enhancement from drug to drug," J. Pharm. Sci. **86**, 1190–1192 (1997).

[12]H. R. Guzman, A. J. McNamara, D. X. Nguyen, and M. R. Prausnitz, "Bioeffects caused by changes in acoustic cavitation bubble density and cell concentration: A unified explanation based on cell-to-bubble ratio and blast radius," Ultrasound Med. Biol. **29**, 1211–1222 (2003).

[13]R. K. Schlicher, H. Radhakrishna, T. P. Tolentino, R. P. Apkarian, V. Zarnitsyn, and M. R. Prausnitz, "Mechanism of intracellular delivery by acoustic cavitation," Ultrasound Med. Biol. **32**, 915–924 (2006).

[14]E. Maione, K. K. Shung, R. J. Meyer, J. W. Hughes, R. E. Newnham, and N. B. Smith, "Transducer design for a portable ultrasound enhanced transdermal drug delivery system," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1430–1436 (2002).

[15]J. F. Tressler, W. Cao, K. Uchino, and R. E. Newnham, "Finite element analysis of the cymbal-type flextensional transducer," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 1363–1369 (1998).

[16]J. Zhang, J. W. Hughes, R. J. Meyer, K. Uchino, and R. E. Newnham, "Cymbal array: A broad band sound projector," Ultrasonics **37**, 523–529 (2001).

[17]R. J. Meyer, A. Dogan, C. Yoon, S. M. Pilgrim, and R. E. Newnham, "Displacement amplification of electroactive materials using the cymbal flextensional transducer," Sens. Actuators A **87**, 157–162 (2001).

[18]N. B. Smith, S. Lee, and K. K. Shung, "Ultrasound-mediated transdermal *in vivo* transport of insulin with low-profile cymbal arrays," Ultrasound Med. Biol. **29**, 1205–1210 (2003).

[19]B. Snyder, S. Lee, R. E. Newnham, and N. B. Smith, "Ferroelectric transducer arrays for transdermal insulin delivery," J. Mater. Sci. **41**, 211–216 (2006).

[20]E. J. Park, J. Werner, and N. B. Smith, "Ultrasound mediated transdermal insulin delivery in pigs using a lightweight transducer," Pharmacol. Res. **24**, 1396–1401 (2007).

[21]M. Pavlovic, K. Wroblewski, Y. Manevich, S. Kim, and J. E. Biaglow, "The importance of choice of anaesthetics in studying radiation effects in the 9L rat glioma," Br. J. Cancer Suppl. **27**, S222–S225 (1996).

[22]N. Kawai, R. F. Keep, and A. L. Betz, "Hyperglycemia and the vascular effects of cerebral ischemia," Stroke **28**, 149–154 (1997).

[23]S. Lee, B. Snyder, R. E. Newnham, and N. B. Smith, "Noninvasive ultrasonic transdermal insulin delivery in rabbits using the light-weight cymbal array," Diabetes Technol. Ther. **6**, 808–815 (2004).

[24]R. E. Newnham, Q. C. Xu, and S. Yoshikawa, "Transformed stress direction acoustic transducer," 4,999,819, 12 March 1991.

[25]R. E. Newnham, Q. C. Xu, and S. Yoshikawa, "Metal-electroactive ceramic composite actuators," 5,276,657, 4 January 1994.

[26]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).

[27]R. E. Newnham and A. Dogan, "Metal-electroactive ceramic composite transducer," 5,729,077, 17 March 1998.

[28] W. P. Mason, "Electrostrictive effect in barium titanate ceramics," Phys. Rev. **74**, 1134–1147 (1948).

[29] A. H. Meitzler, H. M. Obryan, and H. F. Tiersten, "Definition and measurement of radial mode-coupling factors in piezoelectric ceramic materials with large variations in poissons ratio," SU20, 233–239 (1973).

[30] IEEE Standards Board, *IEEE Standard on Piezoelectricity* (Ferroelectrics and Frequency Control Society Standards Committee of the IEEE Ultrasonics, New York, 1988).

[31] N. Setter, *Piezoelectric Materials in Devices* (EPFL Swiss Federal Institute of Technology, Lausanne, Switzerland, 2002).

[32] B. Hamonic, J. C. Debus, J. N. Decarpigy, D. Boucher, and B. Tocquet, "Analysis of a radiating thin-shell sonar transducer using the finite element method," J. Acoust. Soc. Am. **86**, 1245–1253 (1989).

[33] *IEEE Guidè for Medical Ultrasound Field Parameter Measurements* (IEEE, New York, 1990).

[34] AIUM, *Acoustic Output Labeling Standard for Diagnostic Ultrasound Equipment* (American Institute of Ultrasound in Medicine, Laurel, MD, 1998).

[35] N. B. Smith, S. Lee, E. Maione, R. B. Roy, S. McElligott, and K. K. Shung, "Ultrasound-mediated transdermal transport of insulin *in vitro* through human skin using novel transducer designs," Ultrasound Med. Biol. **29**, 311–317 (2003).

[36] S. Lee, V. Nayak, J. Dodds, M. Pishko, and N. B. Smith, "Glucose measurements with sensors and ultrasound," Ultrasound Med. Biol. **31**, 971–977 (2005).

[37] H. Rifkin and D. Porte, *Ellenberg and Rifkin's Diabetes* (Elsevier Science, New York, 1990).

[38] J. E. Shaw *et al.*, "Impaired fasting glucose or impaired glucose tolerance. What best predicts future diabetes in Mauritius?," Diabetes Care **22**, 399–402 (1999).

[39] G. P. Carnevale Schianca, A. Rossi, P. P. Sainaghi, E. Maduli, and E. Bartoli, "The significance of impaired fasting glucose versus impaired glucose tolerance: Importance of insulin secretion and resistance," Diabetes Care **26**, 1333–1337 (2003).

[40] S. Lee, R. E. Newnham, and N. B. Smith, "Short ultrasound exposure times for noninvasive insulin delivery in rats using the lightweight cymbal array," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 176–180 (2004).

Micromachined optical microphone structures with low thermal-mechanical noise levels

Neal A. Hall,[a)] Murat Okandan, and Robert Littrell
Sandia National Laboratories, Albuquerque, New Mexico 87185

Baris Bicen and F. Levent Degertekin
G.W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Micromachined microphones with diffraction-based optical displacement detection have been introduced previously [Hall *et al.*, J. Acoust. Soc. Am. **118**, 3000–3009 (2005)]. The approach has the advantage of providing high displacement detection resolution of the microphone diaphragm independent of device size and capacitance—creating an unconstrained design space for the mechanical structure itself. Micromachined microphone structures with 1.5-mm-diam polysilicon diaphragms and monolithically integrated diffraction grating electrodes are presented in this work with backplate architectures that deviate substantially from traditional perforated plate designs. These structures have been designed for broadband frequency response and low thermal mechanical noise levels. Rigorous experimental characterization indicates a diaphragm displacement detection resolution of 20 fm/$\sqrt{\text{Hz}}$ and a thermal mechanical induced diaphragm displacement noise density of 60 fm/$\sqrt{\text{Hz}}$, corresponding to an A-weighted sound pressure level detection limit of 24 dB(A) for these structures. Measured thermal mechanical displacement noise spectra are in excellent agreement with simulations based on system parameters derived from dynamic frequency response characterization measurements, which show a diaphragm resonance limited bandwidth of approximately 20 kHz. These designs are substantial improvements over initial prototypes presented previously. The high performance-to-size ratio achievable with this technology is expected to have an impact on a variety of instrumentation and hearing applications.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2769615]

PACS number(s): 43.38.Gy, 43.38.Kb, 43.38.Ar, 43.38.Bs [AJZ]   Pages: 2031–2037

I. INTRODUCTION

Micromachined microphones with diffraction-based optical displacement detection have been introduced previously.[1] The approach has the advantage of providing high displacement detection resolution of integrated microphone diaphragm structures independent of device size, capacitance, and mechanical and acoustical design features. This attribute allows for the use of innovative microphone backplate designs with minimal area and low squeeze-film resistance—the dominant damping mechanism and thermal-mechanical noise source in condenser microphones.[2,3] This feature may enable microphones with size characteristic of the smallest microphones demonstrated to date (with the essential semiconductor components contained within approximately 1.5 mm³ volume), but with detection performance characteristic of the highest precision measurement microphones [below 20 dB(A) noise levels]. Single element microphones with this property may be useful for noninvasive instrumentation applications and audio applications where covertness is desired, while microfabricated arrays may be ideally suited for broadband sound intensity measurements and advanced speech recognition applications utilizing ambient noise suppression algorithms designed for miniature acoustic apertures.[4,5]

An embodiment of the sensing technology is summarized in Fig. 1 and consists of a silicon die with a surface micromachined microphone diaphragm and a monolithically integrated rigid diffraction grating back electrode. The diaphragm-grating structure is illuminated with coherent light from a semiconductor laser such as a vertical cavity surface emitting laser (VCSEL) fabricated on a separate die, and the intensities of the diffracted orders that result are monitored with integrated photodetection electronics. As analyzed in detail previously, the integrated grating structure functions as a beam splitter—passing a portion of the incident radiation in between the fingers to the diaphragm and back to accrue additional phase with respect to the light reflecting directly off of the grating fingers.[6] The intensities of the diffracted orders are modulated by the diaphragm displacement with a relationship identical to that of a full-scale Michelson interferometer. The diaphragm and grating electrode can be made electrically conductive, which enables capacitive actuation capabilities of the diaphragm as highlighted schematically with the applied voltage in Fig. 1. We have demonstrated the use of this integrated feature for calibration of the displacement detection scheme in which dc forces are generated to pull the diaphragm through optical

---

a)Electronic mail: nahall@alumni.utexas.net
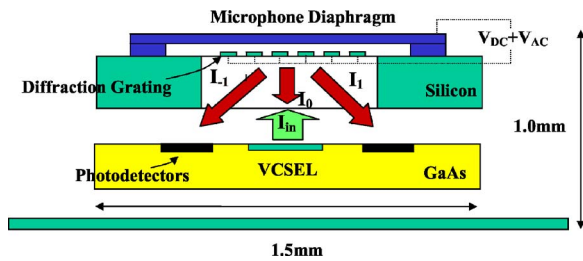
FIG. 1. (Color online) Schematic of the diffraction-based optical microphone technology.

FIG. 2. (Color online) Micrograph of a completed structure from the topside view.

interference cycles, and for broadband dynamic self-characterization in which electrostatic forces are generated from dc to over 100 kHz.[1,7]

Significant developments have been demonstrated with the approach using prototype ultrasonic transducers fabricated on quartz, including the demonstration of phased-array ultrasonic imaging in air, optoelectronic integration of ultrasound sensors in 2.5 mm³ volume using custom-designed complementary metal-oxide semiconductor photodiode arrays, and the detection of sound using pulsing of the light source to demonstrate low-power operation.[8,9] Recently, prototype microphones were fabricated using Sandia National Laboratories' dedicated silicon-based process technology.[10] These structures utilize surface micromachined silicon-nitride diaphragms and bulk-silicon microphone backplates perforated using deep reactive ion etching. A rigorous characterization procedure for these components has been presented, which consists of (1) calibration of the displacement detection sensitivity, (2) dynamic response characterization for experimental identification of key system parameters, and (3) utilization of these parameters to simulate thermal mechanical noise spectra for comparison with experimentally measured data.[1] Displacement of the microphone diaphragm with 2.4 pm rms A-weighted resolution has been demonstrated—illustrating the potential for realizing miniature high performance microphones. Characterization of these initial structures also revealed the potential for several mechanical and acoustical design improvements to the micromachined structure, including a redesigned backplate architecture for reduced damping, enhanced bandwidth, and low thermal mechanical noise. In what follows, optical microphone structures with advanced backplate designs fabricated using Sandia's dedicated silicon-based microfabrication technology are presented and characterized in detail using the above-outlined procedure—demonstrating 1.5-mm-diam microphone structures with 24 dB(A) thermal mechanical noise level and 20 kHz detection bandwidth. These figures constitute a 15 dB and 4× improvement in noise and bandwidth, respectively, as compared to initial prototype designs.[1] The relevance of these measurements are then summarized, and the potential to achieve 15 dB(A) performance from these 1.5-mm-diam structures with relatively simple design changes is discussed.

II. MICROMACHINED OPTICAL MICROPHONE STRUCTURES

The processing sequence used for fabrication of the diffraction-based optical microphone structures is similar to those described in detail previously, but with significant improvements to create the advanced designs presented here.[10] Circular polysilicon diaphragms with 2.25 μm thickness, 1.5 mm diameter, and near-zero residual stress are surface micromachined on a silicon substrate to form the microphone diaphragm as shown schematically in Fig. 1 and in an actual micrograph in Fig. 2, where a completed structure is viewed from the topside. The diaphragm is suspended above a bulk-silicon backplate structure with a separation of 6 μm as determined by the sacrificial oxide thickness used during the fabrication process. The details of the backplate are only visible from back side views of the structure—presented at varying zoom levels in the micrographs of Fig. 3. The backplate consists of three bulk-silicon arm-like structures which hold a surface micromachined grating rigidly in place 6 μm beneath the compliant diaphragm. The grating is etched into a 0.8-μm-thick silicon nitride layer with a 4 μm period. This nitride layer exists above the silicon substrate across its entire surface and provides electrical isolation between the microphone diaphragm and bulk silicon substrate. Although not visible in any views presented here, an electrically conductive 0.3-μm-thick polysilicon layer is patterned above the nitride to form a ground plane electrode with a footprint similar to the three-arm structure itself. This electrode makes electrical contact with the bulk silicon substrate through small etch holes in the nitride. A similar etch is made outside of and away from the circumference of the diaphragm to create a wirebond pad for the ground electrode as can be seen in Fig. 2. The isolated electrodes enable the generation of well-controlled forces through capacitive actuation as demonstrated in the following sections. In a final process step, 100 nm of gold is sputtered on the back side of the structure to enhance the optical reflectivity of the grating and diaphragm.

The film resistance associated with the squeezing of air in the thin 6 μm gap between the diaphragm and backplate is the dominant damping mechanism and thermal-mechanical noise source of the structure. The backplate shown in Fig. 3 has been tailor designed for broadband frequency response and low thermal-mechanical noise using an advanced thin-film damping model[11] able to study complex backplate ge-

(a)

(b)

(c)

FIG. 3. (Color online) Micrographs of diffraction-based optical microphone structures taken from the back side at varying zoom levels showing (a) the three-arm backplate structure and (b) and (c) details of the diffraction grating.

FIG. 4. (Color online) Experimental configuration used for displacement-sensitivity calibration, dynamic characterization, and noise measurements.

bold representation of a critical advantage of the detection approach—the ability to design miniature acoustic sensing structures with minimal constraints related to device capacitance. Neighboring microphone elements on the same chip can also be partially observed in Figs. 2 and 3(a) as these silicon components have been designed for miniature array applications. A final important note regarding the device structure is the use of chemical mechanical polishing of the 6-μm-thick sacrificial oxide layer to remove topography resulting from the underlying grating structure during deposition. This planarization process results in flat diaphragms and nearly ideal optical interference curves as presented in Sec. III.

III. CALIBRATION OF DISPLACEMENT DETECTION SENSITIVITY

The optimal signal read-out architecture for the approach consists of capturing the intensities of the zero and higher order beams as summarized in Fig. 1, as the intensities of all beams are modulated by the diaphragm motion and contain signal. The experimental configuration used in this work for investigative purposes is summarized in Fig. 4, which is more general and allows for characterization of a variety of optics-based sensors under development. Light from a laser is passed through a beamsplitter to create a beam input to the microphone structure and a reference beam which can be used for cancellation of common-mode intensity noise in the light source. The intensities of a first order diffracted beam from the microphone structure and the reference beam are monitored with silicon photodiodes.[13] The reference beam is not needed and is covered for the displacement sensitivity calibration and dynamic characterization measurements. In these cases, the first order photocurrent passes through a current-to-voltage amplifier with 100 kΩ transimpedance gain. Experiments were performed using both an 850 nm single mode VCSEL and a 630 nm laser diode as the light source.

Calibration of the displacement detection sensitivity is demonstrated in Fig. 5(a) in an experiment using the laser diode as the light source. The first-order signal is recorded while pulling the diaphragm toward the grating via the application of electrostatic potential applied across the elec-

ometries that are a substantial departure from traditional perforated plate structures commonly employed in capacitive microphones. The degree of perforation for such structures—which directly affects damping and thermal noise levels[2]—is limited by capacitance requirements.[3,12] The structure presented in Fig. 3, in contrast, consists of mostly open area and therefore very small active capacitance. The images are a

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Hall *et al.*: Micromachined microphones with low thermal noise   2033

(a)

(b)

FIG. 5. (Color online) Experimentally measured optical interference curves using (a) a 630 nm laser diode and (b) an 850 nm VCSEL as the light source.

trodes. Approximately two full interference cycles are traversed in this particular experiment as the bias potential is swept in both the forward and backward directions to illustrate the stability and repeatability of the device structure. The interference curves exhibit nearly ideal modulation efficiency,[14] greater than 95%. High modulation efficiency has also been demonstrated using VCSELs as the light source, as presented in an interference curve measurement in Fig. 5(b). Generally speaking, these curves are slightly different due to different optical wavelengths and power levels. Maximum displacement detection sensitivity and linearity occur at points of maximum slope on the interference curves, such as the point labeled B in Fig. 5(a) corresponding to an applied bias potential of 17.5 V. The displacement detection sensitivity at such points for any scheme with Michelson-type modulation is equal to the peak-to-peak swing in signal level scaled by $2\pi/\lambda$. For the parameters used in this particular experiment and for the peak output voltage signal levels at points A and C in Fig. 5(a), for example, the displacement detection sensitivity can be computed as 100 V/μm. The electrostatic actuation capability also enables the possibility of integrating a self-tuning feature in

(a)

(b)
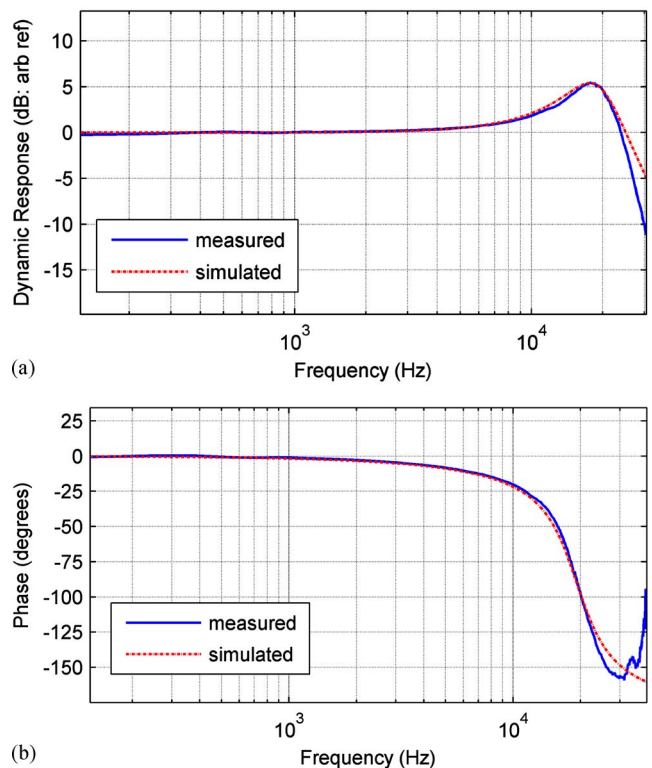
FIG. 6. (Color online) Experimentally measured and simulated (a) amplitude and (b) phase electrostatic frequency response functions of the micromachined microphone structure

which feedback electronics are used to maintain the diaphragm at a position of maximum detection sensitivity in the presence of changing ambient conditions (i.e., temperature, humidity) that would otherwise force the diaphragm to drift.

Calibration of the displacement sensitivity is necessary for converting measured voltage noise spectra at the sensor output into corresponding units of diaphragm displacement as presented in the following sections. The directly measured displacement noise spectra are compared with thermal mechanical noise simulations based on system parameters derived from dynamic characterization.

IV. DYNAMIC CHARACTERIZATION

The dynamic frequency response function for the device relating the center diaphragm displacement to electrostatically applied input pressure has been experimentally measured using an SRS 785 dynamic signal analyzer (DSA), whereby a broadband chirp is electrostatically applied to the device and the fast fourier transform (FFT) coefficients of the measured output signal are normalized by the FFT coefficients of the applied input. The amplitude and phase response are shown in Figs. 6(a) and 6(b), respectively. The device has a flat mechanical response from dc to 10 kHz, and a 5 dB rise at the resonant frequency near 20 kHz.

A rigorous finite-element based dynamic model has been developed for these structures based on modal superposition in which a system of modal equations for the diaphragm structure is derived, and modal damping coefficients are computed using a finite-element based simulation of the squeeze film phenomena. The model can be used to simulate

frequency response functions for arbitrary shaped forcing inputs and thermal mechanical noise spectra across a wide frequency range spanning many modes of the device. An abbreviated form is employed here in which only the first modal equation is retained and the resonant frequency and damping ratio are simply chosen to fit the experimentally measured frequency response data. Simulations using values of 19.25 kHz and 0.28 for the resonant frequency and damping ratio, respectively, are presented in Fig. 6 and match well with the measured frequency response data.[15] These parameters, along with the well-known device dimensions and density of the microfabrication materials, can be used to simulate the thermal-mechanical diaphragm displacement, δ, using the following relation,

$$\delta = \frac{\sqrt{8kT\zeta_1\omega_1}}{\|\omega_1^2 - \omega^2 + 2i\zeta_1\omega_1\omega\|}\Phi_1, \tag{1}$$

where k, T, ζ_1, ω_1, and Φ_1 are Boltzmann's constant, temperature, and the damping ratio, resonant frequency, and mass-normalized modal displacement profile of the fundamental mode. The denominator represents the well-known transfer function of the first modal equation, where only the absolute value is relevant in this form due to the random nature of the noise excitation. Derivation of the expression follows directly from the application of the generalized Nyquist relation to the modal equation of motion.[16] Similar relations have been employed in the analysis and design of precision instruments such as cantilevers used in atomic force microscopy (AFM).[17] The thermal mechanical displacement noise spectrum occurring at the center of the diaphragm structure is simulated and compared with measured noise spectra in Sec. V.

V. NOISE SPECTRA MEASUREMENTS

To enable assessment of the thermal mechanical noise of the structure, ambient acoustic noise is made negligible by enclosing the sensor in a steel cube with a glass window to allow for optical access to the sensor. The ability of the steel enclosure to mitigate ambient room noise below that of the thermal-mechanical noise of the sensor throughout the measurement frequency range is verified by applying controlled sound outside of the enclosure with a speaker at levels well above the ambient room noise while observing that the sensor is indeed immune to the applied sound—a previously employed verification technique used in thermal-mechanical noise measurements.[3]

For noise measurements, excess intensity noise in the laser diode is canceled by equalizing the dc photocurrents in the first-order and reference beams and then subtracting one from the other for common mode intensity noise rejection before passing the resulting photocurrent through a current to voltage amplifier with a 100 kΩ gain. The equalization can be performed in the optical domain by adjusting the intensity in the reference beam using the variable attenuator shown in Fig. 4, or electronically using an autobalanced feedback circuit described by Hobbs.[18] Across the measurement frequency range presented here, both approaches were equally effective at canceling intensity noise of the laser diode to
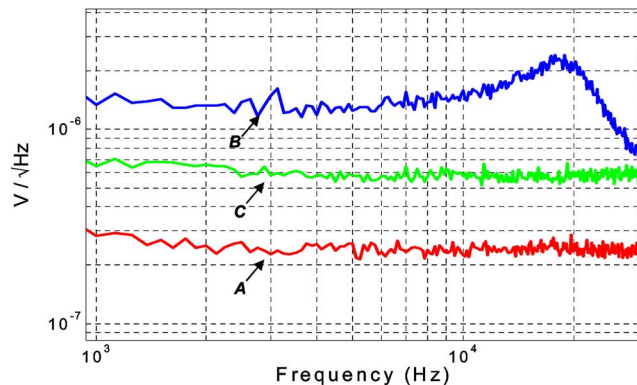
FIG. 7. (Color online) Measured voltage noise spectra at the sensor output at points of zero and maximum displacement sensitivity.

levels below the quantum shot noise of the laser light—the theoretical displacement detection noise limit.

Voltage noise spectra at the output of the amplifier circuit are first recorded[19] with the diaphragm positioned at locations resulting in zero displacement detection sensitivity [i.e., the troughs and peaks of the interference curve where the change in light intensity with respect to small changes in diaphragm displacement is zero, corresponding to points A and C in Fig. 5(a)]. The dc signal voltage levels at points A and C in this particular experiment are 0.404 and 3.45 V, respectively, corresponding to 10 and 86 μW of incident light power on the signal photodiode. Theoretically computed quantum shot noise levels for these conditions are 161 and 470 nV/$\sqrt{\text{Hz}}$,[20] which are in close agreement (within 2.8 and 1.5 dB, respectively) with the measured spectra labeled A and C in Fig. 7—illustrating shot noise limited detection at these locations. In the absence of any diaphragm motion, the noise level measured at a point corresponding to point B of the interference curve would fall in between these levels as shot noise scales with the square root of the dc signal level. The measured noise spectrum at this point, labeled B in Fig. 7, is significantly higher than the shot noise limit and similar in shape to the frequency response of the device as presented in Fig. 6. The observed noise spectrum is dominated by thermal-mechanical displacement of the diaphragm.

Additional confirmation of this conclusion is obtained from close agreement between the measured and simulated thermal noise spectra. Measured voltage noise spectrum B from Fig. 7 is converted to displacement units using the calibrated displacement detection sensitivity and is presented in Fig. 8. The simulation following Eq. (1) is also presented, showing exceptional agreement (within 1.0 dB) with the measured thermal noise spectrum. The measured thermal mechanical displacement noise density throughout the audio frequency range is approximately 60 fm/$\sqrt{\text{Hz}}$, as observed in Fig. 8. This measured displacement noise can be referred to pressure through the compliance of the diaphragm structure, which is inferred as 22 nm/Pa through a procedure described in detail previously where measured resonant data in vacuum are used in conjunction with well-known inertial properties of the diaphragm.[10] Independent and direct measurement of diaphragm compliance using a commercial AFM system with a calibrated cantilever is within 6% agreement. This

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Hall *et al.*: Micromachined microphones with low thermal noise   2035
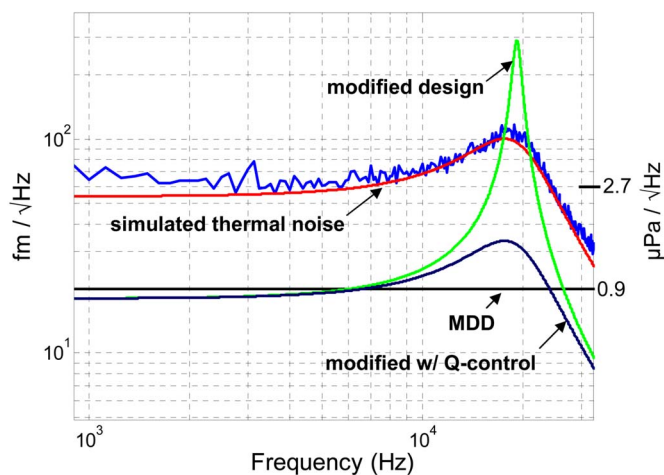
FIG. 8. (Color online) Measured and simulated displacement noise spectra of the existing micromachined optical microphone structure, and simulated thermal-mechanical noise spectra of modified designs.

value results in a thermal-noise limited pressure resolution of 2.7 μPa/$\sqrt{\text{Hz}}$ for the device structure presented in this work. For reference, 2.7 μPa/$\sqrt{\text{Hz}}$ noise level throughout the audio frequency range corresponds to an A-weighted sound pressure level approximately equal to 24 dB(A).

VI. DISCUSSION AND CONCLUSION

It should be noted that claiming the demonstration of a microphone with 24 dB(A) noise level in this work is not technically correct, as the structures presented have not been fully packaged as microphones and noise measurements below 1 kHz have not been presented. Thompson *et al.*, for example, have demonstrated that packaging related enclosures in miniature hearing aid microphones constitute the dominant thermal mechanical noise source across some significant frequency bands—concluding that packaging must be performed with acute attention to acoustic design detail for miniature high-performance sensors.[21] With this said, the thermal mechanical noise levels of the structures investigated here represent a fundamental detection limit 12 dB lower than the only existing commercial Microelectromechancial System (MEMS) microphone of comparable size[22] and approximately equal to those of demonstrated MEMS capacitive instrumentation microphones that have undergone rigorous design optimization.[12]

The shot-noise limited displacement detection levels presented in Fig. 8 have been achieved down to 100 mHz and up to 100 kHz using referenced detection principles.[18] It should be noted that referenced detection capability is inherently built into the fully-integrated sensor shown in Fig. 1, as the zero-order signal beam is complementary to the higher order signal beams. A differential read-out configuration results in both intensity noise cancellation as well as doubling of the displacement sensitivity as previously demonstrated.[9]

The shot noise limited minimum detectable displacement (MDD) level corresponding to diaphragm position B is approximately 20 fm/$\sqrt{\text{Hz}}$ (labeled with a flat solid trace in Fig. 8), which is roughly 3$\times$ lower than the thermal mechanical displacement noise level. The sensor would benefit

from additional reduction in the thermal mechanical displacement noise down to this value, which requires a 9$\times$ reduction in damping owing to the square root in Eq. (1). This can be achieved, for example, by further reducing the backplate area or by increasing the gap height by a factor 2 $\times$ as the backplate damping is known to scale inversely with the cube of the gap height. These are relatively simple design changes. The thermal noise spectrum for such a structure has been simulated and presented as a "modified design" in Fig. 8. Given the diaphragm compliance of 22 nm/Pa, the corresponding pressure noise resolution for this design is 0.91 μPa$\sqrt{\text{Hz}}$, corresponding to 15 dB(A)—a figure comparable to only the highest-fidelity instrumentation microphones in existence today. The undesirably high Q resulting in the response can be eliminated using a force-feedback technique in which internal electrostatic forces are applied to the diaphragm in proportion to its velocity. This technique has been used in AFM measurement systems to critically damp the cantilever response for enhanced imaging speed while preserving the low thermal mechanical noise levels characteristic of the open-loop high-Q resonance of the cantilever, a technique referred to as Q-control.[23–25] A simulation of the modified structure employing this feedback configuration is also presented in Fig. 8.

The modified designs presented in Fig. 8 are optimal in the sense that the displacement detection resolution is nearly equal to the thermal mechanical displacement noise of the diaphragm. Since the largest noise source sets the detection limit, this implies that neither aspect of the system is overdesigned with respect to the other. With capacitive read-out, displacement (or pressure) resolving capability cannot be designed independent of the thermal displacement (or pressure) noise level. Design techniques for lowering thermal noise (reduced backplate area and increased gap height) are precisely responsible for deteriorating the capacitive sensitivity to displacement and, in turn, displacement (or pressure) detection resolution. This fundamental design conflict limits the achievable performance of miniature capacitive microphones.[12] The absence of this conflict in the optical case has enabled rapid mechanical and acoustical design improvements as demonstrated by comparing the structures summarized here to initial prototypes.[1]

Additional improvements leading to microphones with 15 dB(A) performance and 1.5 mm diaphragm size have been projected and may represent a significant breakthrough in noninvasive scientific and industrial measurement applications. Furthermore, the micromachining aspect may be ideal for miniature matched-element array construction, enabling sound intensity measurements as well as the formation of directional microphones with first and higher order directivity patterns for covert speech recognition systems used in noisy environments. Although not presented here, microphone structures with 1 mm diaphragm diameter have been fabricated for broadband applications and show a flat response from dc to 60 kHz. The combination of bandwidth and fidelity may make these structures ideal for acoustic signature recognition applications and industrial diagnostics.

Robust array integration of this technology and demonstration of unique measurement capabilities is the subject of current and future work for the authors.

ACKNOWLEDGMENTS

[1]N. A. Hall, B. Bicen, M. K. Jeelani, W. Lee, S. Qureshi, M. Okandan, and F. L. Degertekin, "Micromachined microphones with diffraction based optical displacement detection," J. Acoust. Soc. Am. **118**, 3000–3009 (2005).

[2]T. B. Gabrielson, "Mechanical-thermal noise in micromachined acoustic and vibration sensors," IEEE Trans. Electron Devices **40**, 903–909 (1993).

[3]V. Tarnow, "The lower limit of detectable sound pressures," J. Acoust. Soc. Am. **82**, 379–381 (1987).

[4]M. Stanacevic, G. Cauwengberghs, and G. Zweig, "Gradient flow adaptive beamforming and signal separation in a miniature microphone array," Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, Orlando, FL, 2002, pp. 416–419.

[5]M. Stanacevic and G. Cauwenberghs, "Micropower gradient flow acoustic localizer," IEEE Trans. Circuits Syst., I: Fundam. Theory Appl. **52**, 2148–2157 (2005).

[6]N. A. Hall and F. L. Degertekin, "Integrated optical interferometric detection method for micromachined capacitive acoustic transducers," Appl. Phys. Lett. **80**, 3859–3861 (2002).

[7]N. A. Hall and F. L. Degertekin, "Self-calibrating micromachined microphones with integrated optical displacement detection," in *Transducers*, Springer-Verlag, Munich, Germany, 2001, pp. 118–121.

[8]N. A. Hall, W. Lee, and F. L. Degertekin, "Capacitive micromachined ultrasonic transducers with diffraction-based integrated optical displacement detection," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 1570–1580 (2003).

[9]W. Lee, N. A. Hall, Z. Zhou, and F. L. Degertekin, "Fabrication and characterization of a micromachined acoustic sensor with integrated optical readout," IEEE J. Sel. Top. Quantum Electron. **10**, 643–651 (2004).

[10]N. A. Hall, M. Okandan, and F. L. Degertekin, "Surface and bulk micromachined optical displacement sensor fabricated with Swift-Lite™," J. Microelectromech. Syst. **15**, 770–776 (2006).

[11]J. Mehner, W. Doetzel, B. Schauwecker, and D. Ostergaard, "Reduced order modeling of fluid structural interactions in MEMS based on model projection techniques," in *Transducers*, 2003, pp. 1840–1843.

[12]P. R. Scheeper, B. Nordstrand, J. O. Gullov, B. Liu, T. Clausen, L. Midjord, and T. Storgaard-Larsen, "A new measurement microphone based on MEMS technology," J. Microelectromech. Syst. **12**, 880–891 (2003).

[13]Advanced Photonix SD 200-11-31-241 type photodiodes were used for all measurements. To eliminate the possibility of additional noise in the system, the windows were removed from the TO cans of these packaged components.

[14]Modulation efficiency is defined as the range traversed in the signal output divided by the peak value in the signal output. For a given light power, the sensitivity, dynamic range, and efficiency of interferometric sensors are directly proportional to modulation efficiency.

[15]It should be noted that similar dynamic characterization experiments in vacuum show several orders of magnitude less damping for the sensor, confirming that squeeze film damping across the backplate—as opposed to structural damping—is indeed the dominant damping mechanism for the structure.

[16]H. B. Callen and T. A. Welton, "Irreversibility and generalized noise," Phys. Rev. **83**, 34–40 (1951).

[17]M. V. Salapaka, H. S. Bergh, J. Lai, A. Majumdar, and E. McFarland, "Multi-mode noise analysis of cantilevers for scanning probe microscopy," J. Appl. Phys. **81**, 2480–2487 (1997).

[18]P. C. D. Hobbs, "Ultrasensitive laser measurements without tears," Appl. Opt. **36**, 903–920 (1997).

[19]All noise measurements are performed using an SR785 dynamic signal analyzer.

[20]The computed shot noise levels account for the incoherent addition of shot noise in the signal and reference beams, resulting in a level $\sqrt{2}$ higher than that contained in a single beam.

[21]S. C. Thompson, J. L. LoPresti, E. M. Ring, H. G. Nepomuceno, J. J. Beard, W. J. Ballad, and E. V. Carlson, "Noise in miniature microphones," J. Acoust. Soc. Am. **111**, 861–866 (2002).

[22]See, for example, Knowles Acoustics product data sheet SPM0103NE3.

[23]K. J. Bruland, J. L. Garbini, W. M. Dougherty, and J. A. Sidles, "Optimal control of ultrasoft cantilevers for force microscopy," J. Appl. Phys. **83**, 3972–3977 (1998).

[24]J. Mertz, O. Marti, and J. Mlynek, "Regulation of a microcantilever response by force feedback," Appl. Phys. Lett. **62**, 2344–2346 (1993).

[25]T. Sulchek, R. Hsieh, J. D. Adams, G. G. Yaralioglu, S. C. Minne, C. F. Quate, J. P. Cleveland, A. Atalar, and D. M. Adderton, "High-speed tapping mode imaging with active Q control for atomic force microscopy," Appl. Phys. Lett. **76**, 1473–1475 (2000).

Validity of the limp model for porous materials: A criterion based on the Biot theory

Olivier Doutres,[a] Nicolas Dauchez, Jean-Michel Génevaux, and Olivier Dazel
Laboratoire d'Acoustique UMR CNRS 6613, Université du Maine, 72095 Le Mans Cedex 9, France

The validity of using the limp model for porous materials is addressed in this paper. The limp model is derived from the poroelastic Biot model assuming that the frame has no bulk stiffness. Being an equivalent fluid model accounting for the motion of the frame, it has fewer limitations than the usual equivalent fluid model assuming a rigid frame. A criterion is proposed to identify the porous materials for which the limp model can be used. It relies on a new parameter, the frame stiffness influence (FSI), based on porous material properties. The critical values of FSI under which the limp model can be used are determined using a one-dimensional analytical modeling for two boundary sets: absorption of a porous layer backed by a rigid wall and radiation of a vibrating plate covered by a porous layer. Compared with other criteria, the criterion associated with FSI provides information in a wider frequency range and can be used for configurations that include vibrating plates. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2769824]

I. INTRODUCTION

In recent years, there has been a strong effort for modeling the vibroacoustic response of multilayer systems containing porous materials. During the past decade, numerical methods such as finite element methods based on the Biot theory[1,2] have been widely developed for automotive and aeronautic industries. Classical methods consider the displacements of the solid and fluid phases as variables,[3,4] respectively $\mathbf{u}^s$ and $\mathbf{u}^f$, or a mixed formulation[5,6] based on the displacement of the solid phase and the interstitial fluid pressure p. These numerical methods allow one to predict the structural and fluid couplings induced by the poroelastic medium without any kinematic or geometrical assumptions. However, for large size finite element models, these methods can require significant computational time.

To overcome this limitation, one can consider that the porous layer behaves like a dissipative fluid. Two porous "one-wave" formulations can be found: (i) the "rigid frame" model assumes that the solid phase remains motionless (Ref. 2, pp. 79–111) and (ii) the "limp" model assumes that the stiffness of the solid phase is zero but takes into account its inertial effects[7–9] (Ref. 8, Chap. 5-1). Because the motion of the solid phase is considered in the limp model, this model has to be preferred for most of the applications as in transports, for example, car, train, aircraft, where the porous layers are bonded on vibrating plates. However, it is valid since the frame "flexibility" of the porous material has little influence on the vibroacoustic response of the system. The aim of this paper is to identify, for a given porous material, the frequency ranges for which the limp model can be used.

The approximation of the limp model was first proposed by Beranek[7] and studied more in detail by Ingard.[8] It was applied to the poroelastic formulation of Zwikker and

Kosten[10] to model soft acoustical blankets in aircraft. More recently, Dazel *et al.*[9] applied the limp approximation to an alternative displacement formulation based on Biot's poroelasticity equations. This simplified model accounts for (i) the dissipation mechanisms induced by the relative motion between the two phases and (ii) the effects of the motion of the solid phase using an inertia correction.

The validity of the limp model compared to the poroelastic model has been investigated by several authors. Beranek[7] proposed a simple criterion to identify the porous materials for which the limp assumption can be applied. This criterion is based on the ratio of the bulk modulus of the fluid phase K_f over the bulk modulus of the solid phase K_s: The limp assumption can be used if $K_f/K_s > 20$. In his book, Ingard[8] investigates the absorption coefficient of a porous layer covered or not by a screen and backed by a rigid wall [as presented Fig. 1(a)]. He pointed out that the limp model is not valid in the low frequency range situated below the structural frame resonance of the porous layer because this region is mainly controlled by the stiffness. In the same way, Göransson[11,12] investigated the validity of the limp model in the case of an aircraft double wall transmission problem and concluded that the limp model can be acceptable since no resonances of the frame are present in the frequency range of interest. The main conclusion of the previous works is that the use of the limp model depends not only on the properties of the porous layer but also on the boundary conditions which are applied to it.

In this paper, a criterion is proposed to identify the porous materials and the associated frequency ranges for which the limp model can be used according to the boundary conditions applied to the layer. The identification process relies on the derivation of a new frequency-dependent parameter, the frame stiffness influence (FSI), based on the properties of the porous material. This parameter is developed from the Biot theory[1,2] and quantifies the intrinsic influence of the solid-borne wave[2] on the displacement of the interstitial

---

[a]Electronic mail: olivier.doutres.etu@univ-lemans.fr
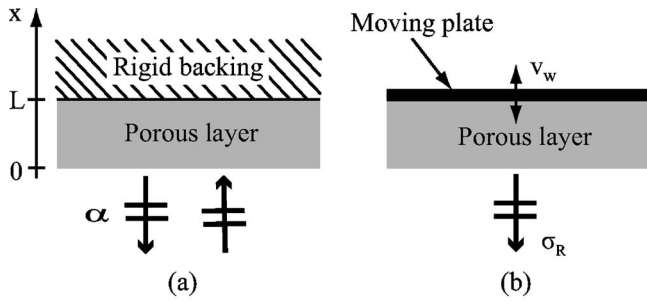
FIG. 1. Two one-dimensional configurations: (a) sound absorption of a porous layer backed by a rigid wall and (b) sound radiation of a porous layer backed by a vibrating wall.

fluid. Critical values of FSI above which the limp model cannot be used are determined for porous materials of thicknesses from 1 to 10 cm and for two boundary condition sets (see Fig. 1): (i) sound absorption of a porous layer backed by a rigid wall and (ii) sound radiation of a porous layer backed by a vibrating wall.

The first part of the paper presents the Biot and the limp models. The FSI parameter, which characterizes the influence of the frame, is then introduced and the method to determine the critical values of FSI is detailed. Critical values are given in Fig. 8 for two boundary condition sets and for ten porous thicknesses. The reader who is mainly interested in the application of the FSI criterion could go directly to the final part of the paper (Sec. VI). In this section, a short explanation on how to use the FSI criterion is proposed and it is illustrated with examples. Finally, a comparison with other criteria validates its efficiency.

II. POROUS MATERIAL MODELING

The limp model introduced in this paper is derived from the Biot theory adapted by Johnson,[13] Allard,[2] and Champoux.[14] This poroelastic model accounts for frame motion, viscous and thermal dissipation mechanisms and is called the Biot model in this paper. In order to achieve a simple analytical description, the modeling is one dimensional, i.e., the porous layer has infinite lateral dimensions in the y and z directions, and only plane waves propagate in the x direction (see Fig. 2). In this case, only two compressional waves are considered.
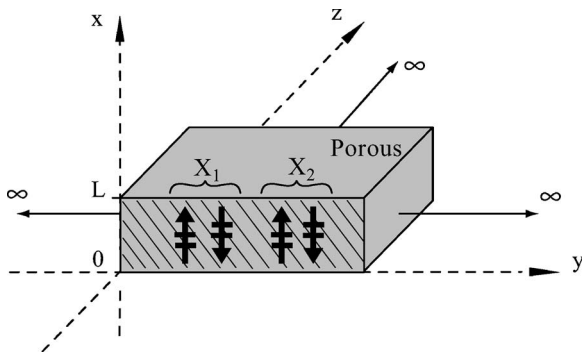
FIG. 2. One-dimensional porous modeling.

A. Biot theory

The motion of the poroelastic medium is described by the macroscopic displacement of solid and fluid phases represented by the vectors $\mathbf{u}^s$ and $\mathbf{u}^f$. Assuming a harmonic time dependence $e^{j\omega t}$, the equation of motion can be written in the following form:[2]

$$-\omega^2 \tilde{\rho}_{11} \mathbf{u}^s - \omega^2 \tilde{\rho}_{12} \mathbf{u}^f = N \boldsymbol{\nabla}^2 \mathbf{u}^s + (\tilde{P} - N) \boldsymbol{\nabla} \boldsymbol{\nabla} \cdot \mathbf{u}^s$$
$$+ \tilde{Q} \boldsymbol{\nabla} \boldsymbol{\nabla} \cdot \mathbf{u}^f, \tag{1}$$

$$-\omega^2 \tilde{\rho}_{12} \mathbf{u}^s - \omega^2 \tilde{\rho}_{22} \mathbf{u}^f = \tilde{Q} \boldsymbol{\nabla} \boldsymbol{\nabla} \cdot \mathbf{u}^s + \tilde{R} \boldsymbol{\nabla} \boldsymbol{\nabla} \cdot \mathbf{u}^f, \tag{2}$$

where the tilde indicates that the associated physical property is complex and frequency dependent. The inertial coefficients $\tilde{\rho}_{11}, \tilde{\rho}_{22}$ are the modified Biot's density of the solid and fluid phases, respectively. The inertial coefficient $\tilde{\rho}_{12}$ accounts for the interaction between inertial forces of the solid and fluid phases together with viscous dissipation. The modified Biot's density can be written in the following form:

$$\tilde{\rho}_{12} = \rho_{12} + j\tilde{b}/\omega, \tag{3}$$

$$\tilde{\rho}_{11} = \rho_{11} - j\tilde{b}/\omega, \tag{4}$$

$$\tilde{\rho}_{22} = \rho_{22} - j\tilde{b}/\omega, \tag{5}$$

with $\rho_{12} = -\phi\rho_f(\alpha_\infty - 1)$, $\rho_{11} = \rho_1 - \rho_{12}$, and $\rho_{22} = \phi\rho_f - \rho_{12}$. ρ_{12} is the added mass due to tortuosity α_∞, ρ_1 the mass density of the porous material, ρ_f the mass density of the interstitial fluid, and ϕ the porosity of the poroelastic material. The coefficient $\tilde{b}$ is related to the viscous and inertial coupling between the two phases. This parameter was first assumed real and constant by Biot.[1] Johnson et al.[13] improved the model by introducing a frequency-dependent complex amplitude to account for the viscous effects. In the present work it is given by

$$\tilde{b} = \phi^2 \sigma \sqrt{1 + j\frac{M}{2}\frac{\alpha_\infty \rho_f}{\phi\sigma}\omega}, \tag{6}$$

$$M = \frac{8\alpha_\infty \mu'}{\sigma\phi\Lambda^2}, \tag{7}$$

where M is the shape factor, μ' is the viscosity of air, σ the resistivity of the porous material, and Λ the characteristic viscous length.

In Eq. (1) and (2), $\tilde{P}$ and $\tilde{R}$ are the bulk modulus of the solid and fluid phases, respectively, and $\tilde{Q}$ quantifies the potential coupling between the two phases. For the majority of sound absorbing materials, these coefficients are[2]

$$\tilde{Q} = (1 - \phi)\tilde{K}_f, \tag{8}$$

$$\tilde{R} = \phi\tilde{K}_f, \tag{9}$$

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Doutres et al.: Limp model criterion   2039

$$\tilde{P} = 2N^* \frac{1-\nu}{1-2\nu} + \frac{\tilde{Q}^2}{\tilde{R}} = \hat{P} + \frac{\tilde{Q}^2}{\tilde{R}}. \qquad (10)$$

The effective bulk modulus of the air in the pores $\tilde{K}_f$ is related to the thermal coupling between the two phases. In the initial paper of Biot,[1] this parameter was real and constant. Champoux and Allard[14] improved the model to account for the thermal effects:

$$\tilde{K}_f = K_a/\tilde{\beta}, \qquad (11)$$

with K_a the adiabatic incompressibility of air and

$$\tilde{\beta} = \gamma - \frac{\gamma - 1}{1 + \left(j\omega \frac{k_0' \rho_f \mathrm{Pr}}{\phi \mu'}\right)^{-1} \tilde{G}'}, \qquad (12)$$

$$\tilde{G}' = \sqrt{1 + j\frac{M'}{2} \frac{k_0' \rho_f \mathrm{Pr}}{\phi \mu'} \omega}, \qquad (13)$$

$$M' = \frac{8k_0'}{\phi \Lambda'^2}. \qquad (14)$$

In Eqs. (11)–(14), γ is the ratio of the specific heats, Pr the Prandtl number, k_0' the thermal permeability, and Λ' the thermal characteristic length. According to Eq. (11), the bulk modulus of the interstitial fluid varies from its isotherm value at low frequencies ($K_f = P_0 = 101.3$ kPa, the atmospheric pressure at 20 °C) to its adiabatic value at high frequencies ($K_f = K_a = \gamma P_0 = 141.2$ kPa at 20 °C).

N^* and ν in Eq. (10) are the complex shear modulus and the Poisson ratio of the frame, considered as an isotropic material. N^* corresponds to the second Lame coefficient μ and can be expressed in terms of the complex Young modulus *in vacuo* E^*:

$$N^* = \frac{E^*}{2(1+\nu)} = \frac{E(1+j\eta)}{2(1+\nu)}, \qquad (15)$$

with η the loss factor of the frame. Note that N^* and E^* are complex and frequency dependent. For the sake of simplicity, these coefficients are assumed to be constant in this work.

Since the model is one dimensional, the displacements are scalars and Eqs. (1) and (2) are rewritten as

$$-\omega^2 \tilde{\rho}_{11} u^s - \omega^2 \tilde{\rho}_{12} u^f = \tilde{P}\nabla^2 u^s + \tilde{Q}\nabla^2 u^f, \qquad (16)$$

$$-\omega^2 \tilde{\rho}_{12} u^s - \omega^2 \tilde{\rho}_{22} u^f = \tilde{Q}\nabla^2 u^s + \tilde{R}\nabla^2 u^f. \qquad (17)$$

Equation (16) can be expressed in a more convenient way to identify the influence of the mechanical properties by introducing the bulk modulus of the solid phase *in vacuo* $\hat{P}$ [see Eq. (10)]. By solving Eq. (16) $-\tilde{Q}/\tilde{R} \times$ Eq. (17), one gets

$$-\omega^2 \frac{\tilde{\rho}_{12}}{\phi} \tilde{\Gamma} u^s - \omega^2 \frac{\tilde{\rho}_{22}}{\phi} \tilde{\gamma} u^f = \hat{P}\nabla^2 u^s, \qquad (18)$$

with

$$\tilde{\Gamma} = \phi\left(\frac{\tilde{\rho}_{11}}{\tilde{\rho}_{12}} - \frac{\tilde{Q}}{\tilde{R}}\right), \quad \tilde{\gamma} = \phi\left(\frac{\tilde{\rho}_{12}}{\tilde{\rho}_{22}} - \frac{\tilde{Q}}{\tilde{R}}\right). \qquad (19)$$

$\tilde{\gamma}$ is a notation introduced by Atalla *et al.*[5] in the mixed displacement-pressure formulation. According to the Biot theory, the two poroelastic equations, Eqs. (17) and (18), can be solved to determine the propagation constants of the two compressional waves. A derivation of the complete system is proposed in Sec. II B.

B. Poroelastic model: A two wave formalism

Compared to classical Biot equations,[1,2] the equations of motion proposed in this paper [Eqs. (17) and (18)] involve the bulk modulus of the frame *in vacuo* $\hat{P}$. The aim of this section is to derive an alternative and a more useful expression of the physical parameters of the Biot waves. This will be of interest for obtaining the FSI.

Using the vector $[\mathbf{w}] = [u^s u^f]^T$, Eqs. (17) and (18) can be reformulated as

$$-\omega^2 [\boldsymbol{\rho}][\mathbf{w}] = [\mathbf{M}]\nabla^2[\mathbf{w}], \qquad (20)$$

with

$$[\boldsymbol{\rho}] = \begin{bmatrix} \tilde{\rho}_{12}\tilde{\Gamma}/\phi & \tilde{\rho}_{22}\tilde{\gamma}/\phi \\ \tilde{\rho}_{12} & \tilde{\rho}_{22} \end{bmatrix}, \quad [\mathbf{M}] = \begin{bmatrix} \hat{P} & 0 \\ \tilde{Q} & \tilde{R} \end{bmatrix}. \qquad (21)$$

Solution of the eigenvalue problem, $-\delta_i^2[\mathbf{w}_i] = \nabla^2[\mathbf{w}_i]$, gives the squared complex wave numbers δ_i^2 of the two compressional waves ($i = 1, 2$),

$$\delta_i^2 = \frac{1}{2}(\mathbf{A}\delta_c^2 + \delta_f^2) + \frac{(-1)^i}{2}\sqrt{(\mathbf{A}\delta_c^2 + \delta_f^2)^2 - 4\mathbf{B}\delta_c^2 \delta_f^2}, \qquad (22)$$

where

$$\mathbf{A} = \left(1 - \frac{\tilde{Q}\tilde{\rho}_{22}\tilde{\gamma}}{\tilde{R}\tilde{\rho}_{12}\tilde{\Gamma}}\right), \quad \text{and} \quad \mathbf{B} = \left(1 - \frac{\tilde{\gamma}}{\tilde{\Gamma}}\right). \qquad (23)$$

The wave numbers of the compressional Biot waves δ_i are written in terms of two characteristic wave numbers: δ_f, the wave number of the rigid frame model described in Sec. II C [see Eq. (30)], and δ_c, the wave number of a wave propagating in a medium that has the bulk modulus of the frame in vacuum and the density of the frame in fluid (it takes into account the inertial and viscous couplings):

$$\delta_c = \omega \sqrt{\frac{\tilde{\rho}_{12}\tilde{\Gamma}}{\phi \hat{P}}} = \omega \sqrt{\frac{\tilde{\rho}_c}{\hat{P}}}, \qquad (24)$$

with

$$\tilde{\rho}_c = \rho_1 - \tilde{\rho}_{12}/\phi. \qquad (25)$$

Note that the expressions of the wave numbers δ_i [Eq. (22)] are equivalent to the classical expressions that can be found in Ref. 2 (p. 130).

The two eigenvectors can be written $[u_1] = [u_1^s u_1^f]^T$, $[u_2] = [u_2^s u_2^f]^T$. Each component of the vectors $[u_i]$ ($i = 1, 2$), corresponds to the displacement of the solid phase u_i^s and the fluid phase u_i^f induced by the propagation of the Biot wave i.

Doutres *et al.*: Limp model criterion

Using Eq. (18), the ratio of the displacement of the frame over the displacement of the air for the two compressional waves is given by

$$\mu_i = \frac{u_i^f}{u_i^s} = \frac{\widetilde{\rho}_{12}\widetilde{\Gamma}}{\widetilde{\rho}_{22}\widetilde{\gamma}}\left(\frac{\delta_i^2}{\delta_c^2} - 1\right). \tag{26}$$

This ratio indicates in which medium the waves mainly propagate. If it is less than 1, the wave is mainly supported by the frame and is referred to as the "frameborne wave." On the contrary, if the ratio is much greater than 1, the corresponding wave is mainly supported by the saturating fluid and is referred to as the "airborne wave." By considering the whole set of porous materials used in this study and presented in Sec. III, one notices that the subscript $i=1$ corresponds to the airborne wave and the subscript $i=2$ corresponds to the frameborne wave.

In the considered geometry (see Fig. 2), the displacement of each phase is due to the propagation of two compressional waves traveling in both directions, and can be written in the following form:

$$u^s(x) = X_1 + X_2, \tag{27}$$

$$u^f(x) = \mu_1 X_1 + \mu_2 X_2, \tag{28}$$

where $X_i = S_i \cos(\delta_i x) + D_i \sin(\delta_i x)$ is the contribution of each compressional wave $i=1,2$, S_i and D_i being set by the boundary conditions.

C. Equivalent fluid models: A one wave formalism

Here, after a brief description of the classical rigid frame model, the limp model is presented. Both models are derived from the Biot theory presented in Sec. II B.

The rigid frame model assumes that the displacement u^s is zero. This assumption applied in Eq. (17) gives a propagation equation on u^f,

$$\widetilde{K}_f \nabla^2 u^f + \omega^2 \frac{\widetilde{\rho}_{22}}{\phi} u^f = 0. \tag{29}$$

The compressional wave, solution of the propagation equation, is characterized by the wave number δ_f,

$$\delta_f = \omega \sqrt{\frac{\widetilde{\rho}_f}{\widetilde{K}_f}}, \tag{30}$$

where $\widetilde{\rho}_f = \widetilde{\rho}_{22}/\phi$.

The limp model is based on the assumption that the frame has no bulk stiffness: $\hat{P}=0$. It is likely associated with "soft" materials like cotton and glass wool. Hence, by considering this assumption in Eq. (18), one gets a simple relation between the displacements of both solid and fluid phases:

$$u^s = -\frac{\widetilde{\rho}_{22}}{\widetilde{\rho}_{12}}\frac{\widetilde{\gamma}}{\widetilde{\Gamma}} u^f. \tag{31}$$

Then, substituting the solid displacement in Eq. (17) by Eq. (31) gives the propagation equation on u^f,

$$\widetilde{K}_f \nabla^2 u^f + \omega^2 \widetilde{\rho}_{\text{limp}} u^f = 0, \tag{32}$$

where

$$\widetilde{\rho}_{\text{limp}} = \frac{\mathbf{B}}{\mathbf{A}}\widetilde{\rho}_f, \tag{33}$$

with $\mathbf{A}$ and $\mathbf{B}$ the two coefficients defined in Eq. (23).

The compressional wave, solution of the propagation equation, is characterized by the wave number $\delta_{\text{limp}} = \omega\sqrt{\widetilde{\rho}_{\text{limp}}/\widetilde{K}_f}$. According to the expression of the density $\widetilde{\rho}_{\text{limp}}$, the wave number of the limp model can be expressed in terms of the wave number of the rigid frame model as

$$\delta_{\text{limp}}^2 = \frac{\mathbf{B}}{\mathbf{A}}\delta_f^2. \tag{34}$$

As in the rigid frame model, the bulk modulus of the limp model is the modified bulk modulus of the saturating fluid $\widetilde{K}_f$. However, the limp model takes into account the effect of the solid phase displacement by its modified effective density $\widetilde{\rho}_{\text{limp}}$, which transcribes the inertia of the solid phase and its interaction with the fluid phase. This property enables use of the limp model for porous material mounted on a vibrating structure as long as the frameborne wave has no influence on the vibroacoustic behavior of the system. In this way, it is less restrictive than the rigid frame model.

III. FRAME STIFFNESS INFLUENCE

The aim of the following is to propose a parameter based on the properties of the porous material that quantifies the influence of the frame stiffness on the porous behavior. This parameter is called FSI.

A. FSI development

The use of the limp model is possible when the contribution of the frameborne wave is negligible in the considered application. This approximation implies the following in the expressions of the solid and fluid displacements [Eqs. (27) and (28)]:

(i) The contribution of the airborne wave X_1 is great compared to the contribution of the frameborne wave X_2; this condition depends mainly on the boundary conditions: Two configurations will be studied in Sec. IV to set critical values of the FSI parameter.

(ii) Considering the fluid motion [Eq. (28)], the displacement ratio μ_1 associated with the airborne wave is great compared to the displacement ratio μ_2 associated with the frameborne wave: $\mu_2/\mu_1 \ll 1$; this condition is independent from the boundary conditions and will be used to build the FSI parameter.

According to Eq. (26), the ratio μ_2/μ_1 can be written in terms of the wave numbers of the Biot waves δ_i [Eq. (22)] and the characteristic wave propagating mostly in the frame δ_c [Eq. (24)]:

$$\frac{\mu_2}{\mu_1} = \frac{(\delta_2^2/\delta_c^2) - 1}{(\delta_1^2/\delta_c^2) - 1}. \tag{35}$$

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Doutres *et al.*: Limp model criterion   2041

From the study of the behavior of 259 porous materials presented in Sec. IV B, the condition $\mu_2/\mu_1 \ll 1$ is met when the following conditions are observed:

(i) $r_1 = \delta_1^2/\delta_c^2$ tends to 0: This occurs when the wave number δ_1 of the airborne wave is small compared to the wave number δ_c.

(ii) $r_2 = \delta_2^2/\delta_c^2$ tends to 1: This occurs when the wave number δ_2 of the frameborne wave is close to the wave number δ_c.

Both r_1 and r_2 ratios exhibit an asymptotic behavior as the frame stiffness becomes smaller: Their convergence is controlled by a common parameter presented in Sec. III B.

B. A simplified parameter

Let us first consider the asymptotic behavior of the ratio r_1. Substituting δ_1 in the expression of r_1 for Eq. (22) gives

$$r_1 = \frac{\delta_1^2}{\delta_c^2} = \frac{1}{2}(\delta_f^2/\delta_c^2 + \mathbf{A})[1 - \sqrt{1-\varepsilon}], \tag{36}$$

with

$$\varepsilon = \frac{4\mathbf{B}\,\delta_f^2/\delta_c^2}{(\delta_f^2/\delta_c^2 + \mathbf{A})^2}. \tag{37}$$

This expression shows that r_1 may tend to zero when ε tends to zero. Hence, by using the first-order Taylor series expansion of the square root function $\sqrt{1-\varepsilon} = 1 - \varepsilon/2$, it becomes

$$r_1 = \frac{(\mathbf{B}\delta_f^2)/(\mathbf{A}\delta_c^2)}{1 + \delta_f^2/(\mathbf{A}\delta_c^2)}. \tag{38}$$

By introducing the wave number of the limp model using Eq. (34), one gets

$$r_1 = \frac{\delta_{\mathrm{limp}}^2/\delta_c^2}{1 + \delta_{\mathrm{limp}}^2/(\mathbf{B}\delta_c^2)}. \tag{39}$$

Thus, r_1 tends to zero when the wave number ratio $\delta_{\mathrm{limp}}^2/\delta_c^2$ tends to zero or when the term $\delta_{\mathrm{limp}}^2/(\mathbf{B}\delta_c^2)$ is much greater than 1. The first condition is used as a parameter characterizing the influence of the frame, denoted the frame stiffness influence,

$$\mathrm{FSI} = \frac{\delta_{\mathrm{limp}}^2}{\delta_c^2} = \frac{\widetilde{\rho}_{\mathrm{limp}}}{\widetilde{\rho}_c}\,\frac{\hat{P}}{\widetilde{K}_f}, \tag{40}$$

with $\widetilde{\rho}_{\mathrm{limp}}$, $\widetilde{\rho}_c$, $\widetilde{K}_f$, and $\hat{P}$ given in Eqs. (33), (25), (11), and (10), respectively. As the Beranek criterion,[7] the characteristic parameter FSI is expressed according to the ratio between the bulk moduli of the two phases. However, in the expression of FSI, each bulk modulus is divided by the density of their corresponding characteristic wave. Thus, Eq. (40) shows that FSI is frequency dependent and it decreases with the bulk modulus of the frame *in vaccuo* $\hat{P}$. Physically, it can be noticed that FSI is small when δ_c is great compared to δ_{limp}, meaning that the interaction between the two associated waves is weak.
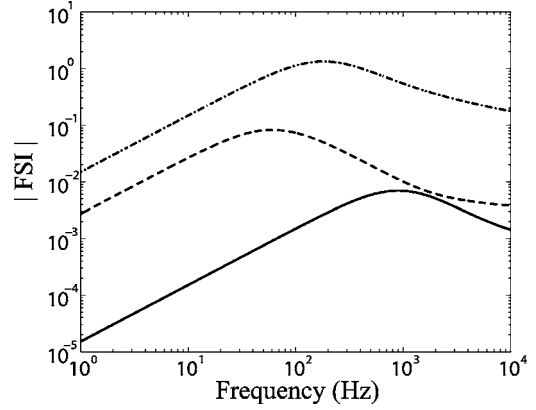
FIG. 3. Evolution of the parameter FSI as a function of the frequency: (—) material A, (---) material B, (-·-) material C, (see Table I).

In the same way, using Eq. (22) and applying the first-order Taylor series expansion of the square root function, the ratio r_2 can be expressed in terms of the parameter FSI as

$$r_2 = \mathbf{A}(1 + \mathrm{FSI}/\mathbf{B}) - \frac{\mathrm{FSI}}{1 + \mathrm{FSI}/\mathbf{B}}. \tag{41}$$

According to Eq. (41), when the parameter FSI tends to zero, r_2 tends to the value of the coefficient $\mathbf{A}$, which is close to 1 for all kinds of porous materials. Hence, the parameter FSI allows one to describe the asymptotic behavior of both r_1 and r_2 and thus the behavior of the ratio μ_2/μ_1. In the following, it is used to identify the porous materials according to the influence of their frame stiffness.

Figure 3 presents the FSI for three characteristic materials (see Sec. IV B and Table I). It is shown that the parameter FSI is frequency dependent and that the material *A*, which has the lower bulk modulus, has the smaller FSI.

IV. DESCRIPTION OF THE CONFIGURATIONS

According to the previous analysis, the limp model can be applied instead of the Biot model if the parameter FSI tends to zero and if the amplitude of the frameborne wave is negligible compared to the amplitude of the airborne wave. This last condition depends primarily on the boundary conditions. In this paper, two different sets of boundary conditions representative of classical applications are investigated. For each configuration, the critical values of FSI under which the limp model can be used are determined from the

TABLE I. Measured properties of materials A, B, and C.

Porous	A	B	C
Airflow resistivity: σ (kN s/m^4)	105	23	57
Porosity: ϕ	0.95	0.95	0.97
Tortuosity: α_∞	1	1	1.54
Viscous length: Λ (μm)	35.1	54.1	24.6
Thermal length: Λ' (μm)	105.3	162.3	73.8
Frame density: ρ_1 (kg/m^3)	17	58	46
Young's modulus at 5 Hz: E (KPa)	1.4	17	214
Structural loss factor at 5 Hz: η	0.1	0.1	0.115
Poisson's ratio: ν	0	0	0.3
Beranek criterion at 5 Hz: $|\widetilde{K}_f/\hat{P}|$	70	6	0.4

response of a wide variety of porous materials. The following presents the configurations and the tested porous materials.

A. The characteristic configurations

Figure 1 presents the two configurations used to evaluate the difference between the Biot and the limp models. The chosen configurations are characteristic of the applications of the porous materials in industrial fields.

In configuration (a), the porous layer is attached to a rigid and motionless backing at $x=L$ and is hit by an acoustic plane wave at $x=0$. The normal-incidence absorption coefficient α of the porous material is derived from the complex reflection coefficient R_x:

$$\alpha = 1 - |R_x|^2, \quad R_x = \frac{Z_n - \rho_f c_f}{Z_n + \rho_f c_f}, \tag{42}$$

with c_f the speed of sound in air and Z_n the normal-incidence surface impedance. This impedance is given by the pressure to the normal velocity ratio in the free air close to the front face of the material; $Z_n = p(0)/v(0)$. Both pressure and velocity are determined using the appropriate boundary conditions: continuity of the normal stress in the solid and fluid phases with the external pressure at the fluid-porous interface $x=0$,

$$\sigma_x^s(0) = -(1-\phi)p(0), \tag{43}$$

$$\sigma_x^f(0) = -p(0), \tag{44}$$

and continuity of the total flow

$$j\omega[(1-\phi)u^s(0) + \phi u^f(0)] = v(0), \tag{45}$$

with u^s and u^f given by Eqs. (27) and (28). On the surface of the porous layer in contact with the backing ($x=L$), the displacement of the air and the frame are equal to zero,

$$u^s(L) = u^f(L) = 0. \tag{46}$$

An analytical formulation of α can be found in Ref. 2 (pp. 21, 137). This coefficient is usually measured in an impedance tube[15] and is used in building or automotive applications.[16]

In configuration (b), the porous layer is excited by a vibrating plate at $x=L$ and radiates in a infinite half-space at $x=0$. This configuration corresponds to trim panels, cars roofs, or airplane floors. The radiation efficiency factor σ_R,[17] defined as the ratio of the acoustic power radiated Π_a over the vibratory power of the piston Π_v, is used as vibroacoustic response:

$$\sigma_R = \frac{\Pi_a}{\Pi_v} = \frac{p(0)v^*(0)}{\rho_f c_f v_w^2}. \tag{47}$$

A vibrating surface area of 1 m² is considered here. Boundary conditions associated with this configuration are:[18] continuity of stress and total flow at $x=0$ and Eqs. (43)–(45) are used. At $x=L$, the velocity of the fluid and the velocity of the frame are both equal to the wall velocity

$$j\omega u^s(L) = j\omega u^f(L) = v_w. \tag{48}$$

TABLE II. Range of values for the properties of the 256 simulated materials.

Airflow resistivity: σ (kN s/m⁴)	1–100
Porosity: ϕ	0.97
Tortuosity: α_∞	1–2
Shape factor: M	1
Viscous length: Λ (μm)	$\sqrt{8\alpha_\infty \mu'/\sigma\phi M}$
Thermal length: Λ' (μm)	3Λ
Frame density: ρ_1 (kg/m³)	10–90
Young's modulus at 5 Hz: E (KPa)	3–200
Structural loss factor at 5 Hz: η	0.01–0.2
Poisson's ratio: ν	0.3

For each configuration, the vibroacoustic response is derived using the transfer matrix method.[2,19] This method assumes the multilayer has infinite lateral dimensions and uses a representation of plane wave propagation in different media in terms of transfer matrices. To ensure a one-dimensional representation, the multilayer is excited by plane waves with normal incidence. The porous layer is either simulated using the Biot model presented in Sec. II B or the limp model presented in Sec. II C.

Note that the two characteristic configurations mainly differ by the kind of excitation applied to the frame: The frame will be less excited by an air domain than by a solid layer. Consequently, the limp model has a greater chance of being suitable for configuration (a) than for configuration (b).

B. Material properties

In order to study the behavior of a wide variety of porous materials, a set of 256 simulated materials and 3 real materials is used. The properties of the real materials presented in Table I have been measured at our laboratory. The viscoelastic characteristics of the porous frames, Youngs modulus E, and loss factor η, were measured at low frequency using the quasistatic method.[20,21]

Material A is a low density glass wool material with a very high airflow resistivity. This material is found in aerospace applications for thermal and sound insulation. Material B is a high density fibrous material and material C is a plastic foam with a stiff skeleton and a high airflow resistivity. Both materials B and C are found in automotive applications, for roof and floor treatment. According to the Beranek criterion (see last line in Table I), only the material A can be considered as a limp material ($|\tilde{K}_f/\hat{P}|=70\gg 20$).

The 256 simulated materials are obtained by setting a random value for the main properties of a porous material: E and η, respectively, the Young modulus and the loss factor of the frame in vacuum, ρ_1 the density of the porous material, σ the airflow resistivity, and α_∞ the tortuosity. The range of values for the properties of the 256 porous materials are presented in Table II. The viscous characteristic length Λ is derived from the shape factor given by Eq. (7). For most of the sound absorbing materials, the shape factor lies between 0.1 and 10 (Ref. 13) and it is chosen here equal to 1. The thermal characteristic length Λ' is three times the viscous characteristic length Λ. For all the simulated materials, the
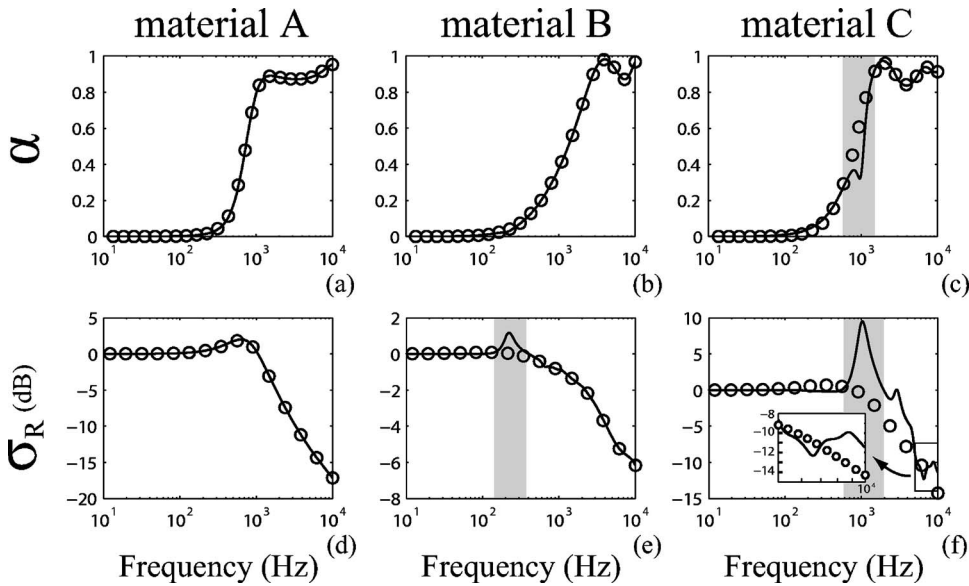
FIG. 4. Vibroacoustic response of the two configurations including the three real materials A, B, and C of thickness 2 cm: (first line) absorption coefficient α, (second line) radiation efficiency σ_R: (—) Biot model, (◯) limp model.

porosity is set to 0.97 and the Poisson ratio ν to 0.3. The set of the randomly simulated materials is built to represent the variety of acoustical materials found in industrial applications.

V. DETERMINATION OF CRITICAL FSI VALUES

In Sec. III B the parameter FSI based on the physical properties of the material has been introduced. The next step is to identify, for each configuration, the critical values of FSI under which the limp model can be used instead of the Biot model. The method to determine a critical value is demonstrated in the case of the absorption coefficient of materials A, B, and C with a thickness of 2 cm. Then, in order to get a critical value independent of the chosen material, the method is generalized with a set of 256 simulated materials, which properties are randomly determined among limits given in Table II. Results for the two configurations and thicknesses from 1 to 10 cm are finally presented.

A. Analysis on materials A, B, and C

1. Three characteristic frequency ranges

Because the influence of the frame is frequency dependent, the study of the critical values of FSI will be carried out in three characteristic frequency bands centered around the first $\lambda/4$ resonance frequency of the frameborne wave, f_r. This frequency can be estimated by simply considering the properties of the frame *in vacuo*:[22]

$$f_r \approx \frac{1}{4L} \sqrt{\frac{\hat{P}}{\rho_1}} = \frac{1}{4L} \sqrt{\frac{E \dfrac{(1-\nu)}{(1+\nu)(1-2\nu)}}{\rho_1}}, \quad (49)$$

where L is the thickness of the porous layer. In the vicinity of f_r, the frame stiffness can have a great influence on both the absorption coefficient[16,22] and the radiation efficiency.[18]

This is illustrated in Fig. 4 (gray zones) where the Biot and the limp simulations of both configurations are presented for materials A, B, and C of thickness 2 cm. In the case of the absorption configuration, a decrease of the absorption

coefficient is observed at the frequency f_r around 1000 Hz for material C [see Fig. 4(c)]. In the case of the radiation configuration, an increase of the radiation efficiency is observed around 200 Hz for material B [see Fig. 4(e)] and 1000 Hz for material C [see Fig. 4(f)]. Higher orders modes $(3\lambda/4, \dots)$ also have an influence in the case of the radiation of material C [Fig. 4(f)].

In the following, three characteristic zones are chosen according to the frequency f_r [Eq. (49)]: $[1, f_r/2]$, the low frequency range (LF); $[f_r/2, 2f_r]$, the midfrequency range (MF); and $[2f_r, 10^4]$, the high frequency range (HF).

2. Estimation of the difference between the models

The difference between the limp and the Biot models is given in third octave bands by the absolute value of the difference between the two vibroacoustic responses V (V being either α or σ_R): $\Delta V = |V_{\text{Biot}} - V_{\text{Limp}}|$. The frequency band $1-10\,000$ Hz is divided in 41 third octave bands. The maximum accepted difference between the two models is set to 0.1 for the absorption coefficient and to 3 dB for the radiation efficiency. These values correspond to a classical industrial demand.

3. Method to determine the FSI critical values

The objective of this section is to determine, according to the configuration and the frequency domain, the maximum value of the FSI under which the limp model can be used. This maximum value is called "critical value."

The method to obtain a critical value is presented for the absorption configuration with materials A, B, and C having a thickness of 2 cm. The difference between the two simulations $|\alpha_{\text{Biot}} - \alpha_{\text{Limp}}|$ is presented in Fig. 5 as a function of FSI in the three frequency ranges for each material. The higher value of FSI above which the maximum acceptable difference between the two models (here 0.1) is exceeded, is found in the medium frequency range at FSI=0.5 [vertical dotted line on Fig. 5(b)]: This value is set as critical value. No critical values are found in the low and the high frequency ranges.
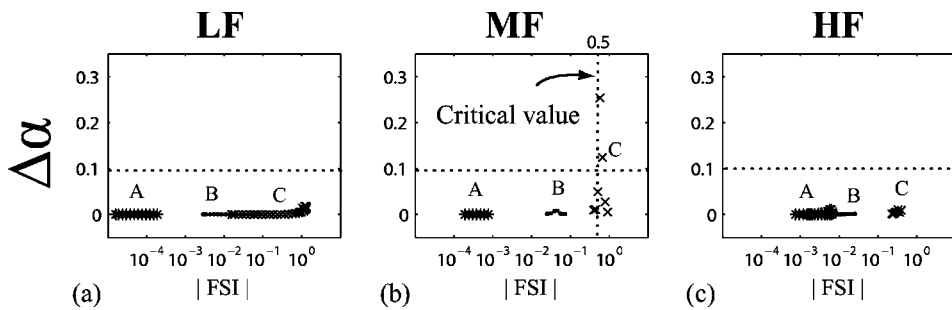
FIG. 5. Difference between the Biot and limp models on the absorption coefficient α of 2-cm-thick materials A ($*$), B ($\bullet$), and C ($\times$), functions of their FSI parameter in the three frequency domains: (a) low frequency band, (b) medium frequency band, and (c) high frequency band.

B. Generalization using 256 simulated porous materials

To establish critical values of FSI that are independent of the chosen material, the preceding study is generalized by using the 256 simulated materials (see Table II). Figures 6(a)–6(c) are equivalent to Fig. 5, but present the difference between the limp and the Biot absorption coefficient derived for 256 simulated materials having a thickness of 2 cm. Figures 6(d)–6(f) present the same results in the case of the radiation configuration.

The first noticeable tendency is that the value of the difference between the limp and the Biot models increases when the value of FSI increases. This is in agreement with the parameter behavior, which increases with the bulk modulus of the frame (see Sec. III B). As stated previously, the critical value of FSI is set when the first parametrized curve exceeds the maximum accepted difference. However, this method can be inappropriate in this case because the limit can depend slightly on the random set of simulated materials. Thus, a simple statistic indicator called the "confidence rate" is used to set the critical values. The FSI range is first divided into 10 bands from 10^{-4} to 10^{0} in a log scale. In each FSI band, the confidence rate gives the number of points below the acceptable limit (0.1 or 3 dB) over the number of points present in this FSI range. This rate, given in percent (solid gray line in Fig. 6), indicates the reliability of using the limp model according to the value of FSI.

The confidence rate of 95% is chosen as a threshold to

set the critical value of FSI, called FSI_{95}. For a FSI larger than FSI_{95}, it is considered that the vibroacoustic response may become notably sensitive to the frame stiffness for most of the porous materials and the Biot model should be preferred. Note that the number of FSI bands used to derive the confidence rate is sufficiently important so that the critical value FSI_{95} does not depend on this number.

In the case of the absorption coefficient in the medium frequency range [see Fig. 6(b)], the critical value is FSI_{95} $=0.16$. No critical values are set in the low and high frequency ranges: The limp model can be used to predict the absorption coefficient for all the acoustic materials. These results obtained for a thickness of 2 cm are summarized in an abacus presented in Fig. 7(a). In this figure, a white cell corresponds to the FSI values for which the limp model can be used. A gray cell corresponds to the FSI values for which the Biot model should be preferred.

Let us consider now the radiation configuration. Figures 6(d)–6(f) present the difference between the two models as a function of FSI, and the obtained critical FSI are summarized in Fig. 7(b). As shown Fig. 6(d), no critical FSI values are set in the low frequency range: The radiation efficiency of a covered plate can be predicted using the limp model for all the acoustic materials for frequencies at less than two times below the first resonance frequency of the frame f_r. Critical values of FSI are set in the medium and the high frequency ranges as seen in Figs. 6(e) and 6(f). The difference between the Biot and the limp models can be great in the high fre-
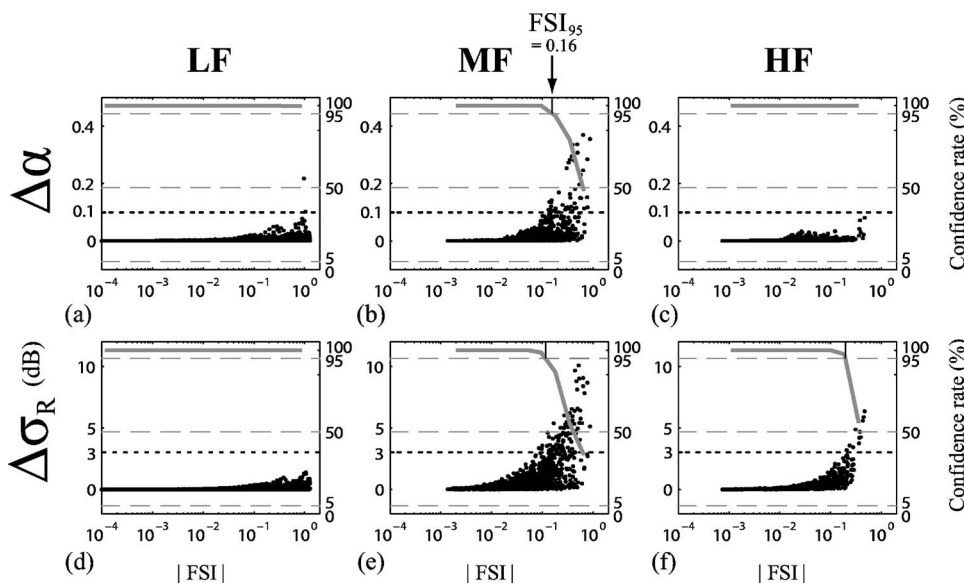
FIG. 6. Evolution of the difference between the Biot and limp models as a function of FSI for the two configurations including the 256 simulated materials of thickness 2 cm: (a)–(c) absorption coefficient, (d)–(f) radiation efficiency of a covered plate; in the three frequency ranges: (first column) low frequency band LF, (second column) medium frequency band MF, and (third column) high frequency band HF.

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Doutres *et al.*: Limp model criterion   2045
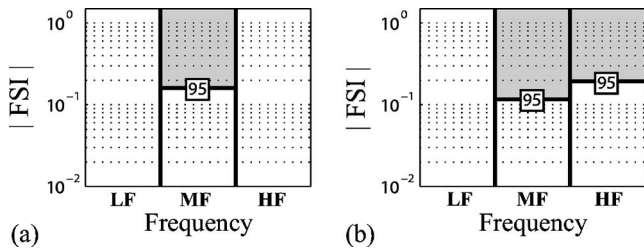
FIG. 7. Critical values of FSI in the three frequency ranges for the two configurations including a porous layer of thickness 2 cm: (a) absorption coefficient, (b) radiation efficiency of a covered plate; white area: The FSI is below the critical value FSI$_{95}$, light gray area: The FSI is above the critical value FSI$_{95}$.

C. Critical values as a function of the thickness

The study of the critical values of FSI is carried out for thicknesses from 1 to 10 cm. Figure 8 presents, in the three frequency domains, the evolution of the critical values of FSI as a function of the layer thickness for both configurations.

In the low frequency range [Figs. 8(a) and 8(d)], most of the configurations do not exhibit a critical FSI. In this frequency range, the wavelength of the frameborne wave is large compared to the thickness of the layer and thus the frame thickness tends to be constant.

In the medium frequency range [Figs. 8(b) and 8(e)], centered around the quarter wavelength resonance, it is shown that the critical values of FSI slightly decrease when the porous thickness increases. This tendency is more pronounced in the high frequency range as shown Figs. 8(c) and 8(f). For example, the absorption coefficient of a material having a FSI of 10^0 can be simulated with the limp model only for thickness up to 4 cm [see Fig. 8(c)]. This shows that the difference between the Biot and the limp models increases with the thickness.

VI. DISCUSSION

It is possible to explain how the FSI criterion should be used. The prediction of the frequency bands for which the limp model can be used is proposed in the case of the absorption coefficient of materials A, B, and C (see Table I) of thickness 2 cm. This method only involves the calculation of the FSI parameter [Eq. (40)] of the tested material and avoids solution of the absorption or radiation problems using the full Biot model and limp model (see Sec. IV A). Finally, the FSI criterion is compared to the Beranek criterion and, in the particular case of the absorption configuration, with classical "rigid frame" criteria.

A. The evaluation procedure and application

The procedure for determining the frequency ranges in which the limp model is valid is as follows:

(i)   The properties of the porous materials have to be measured (see Table I for materials A, B, and C),

(ii) The FSI is derived using Eq. (40); the various parameters ($\hat{P}$, $\tilde{\rho}_c$, and $\tilde{\rho}_{\text{limp}}$) are evaluated using Eqs. (10), (25), and (33); FSI curves are drawn for materials A, B, and C in Figs. 9(a)–9(c), respectively.

(iii) The frequency bands characteristic of the porous behavior are estimated from the center frequency of the medium frequency band f_r [Eq. (49)]; for a 2-cm-thick layer, the frame resonance is $f_{rA} \approx 113$ Hz for material A, $f_{rB} \approx 214$ Hz for material B, and $f_{rC} \approx 990$ Hz for material C.

(iii) The critical values of FSI are chosen in Fig. 8 according to the studied configuration, the thickness of the porous layer, and the frequency band. The data corresponding to the absorption configuration with a layer having a thickness of 2 cm are added on the FSI curves presented in Figs. 9(a)–9(c).

The comparison between the FSI curve and the critical value FSI$_{95}$ enables identification of the frequency bands where the limp model can be used. The results of the comparison for the two configurations are summarized in Figs.

quency range because higher order resonances $(3\lambda/4, 5\lambda/4, \ldots)$ can be excited in the frame thickness and have a noticeable influence on the radiation efficiency.

FIG. 8. Evolution of the FSI critical values as function of the thickness for the two configurations [(a)–(c) absorption coefficient, ((d)–(f) radiation efficiency] in the three frequency ranges: (a), (d) low frequency band LF, (b), (e) medium frequency band MF, (c), (f) high frequency band HF. White zone: limp model valid; grayed zone: limp model not valid.

FIG. 9. (a)–(c) FSI curves and critical values obtained in the case the absorption configuration for a porous layer 2 cm thick: (a) material A, (b) material B, (c) material C; (d)–(f) FSI compared to the critical values for the two characteristic configurations in the three frequency domains: (d) material A, (e) material B, and (f) material C.

9(d)–9(f). It is shown that materials A and B can be simulated using the limp model for both the absorption and the radiation configurations over the whole frequency range (FSI < FSI$_{95}$). These predictions agree with the simulations presented in Figs. 4(a), 4(b), 4(d), and 4(e). Note that the increase of the radiation efficiency of material B [see Fig. 4(e)] induced by the frame resonance does not exceed the maximum accepted difference of 3 dB. For material C, it is predicted that the Biot model should be used in the medium frequency range for both configurations (FSI > FSI$_{95}$). This is confirmed in Figs. 4(c) and 4(f). For the radiation configuration, the use of the Biot model is recommended up to 8 kHz [see Fig. 9(f)]. For higher frequencies, the limp model can be used. This prediction is validated in Fig. 4(f) where the difference between the two simulations is great around the $3\lambda/4$ frame resonance (from 2 to 5 kHz) and does not exceed 3 dB above 8 kHz.

B. Comparison with other criteria

According to the Beranek criterion,[7] only the material A can be simulated with the limp model (for material A, $\widetilde{K}_f/\hat{P}=70 \gg 20$, see Table I). The FSI prediction for this material agrees with Beranek prediction: The absorption coefficient and the radiation efficiency can be simulated with the limp model in the whole frequency range. For the two other materials, the criterion proposed by Beranek is too restrictive: The use of the limp model is banned for both materials B and C ($\widetilde{K}_f/\hat{P}=6$ for material B and 0.4 for material C).

In the particular case of the absorption configuration, the material is generally modeled using the rigid frame assumption (see Sec. II C). To ensure the validity of this assumption, one can use a frequency criterion proposed by Biot.[1] This criterion gives a high critical frequency limit above which the viscous forces only have a noticeable effect in a layer of air in contact with the frame, of equal thickness to the viscous skin depth: The frame cannot be excited by an acoustical wave and can be considered rigid. This frequency is expressed as the ratio of the airflow resistivity of the porous

material over the density of the interstitial fluid: $f_{\mathrm{Biot}} = \sigma/2\pi\rho_f$. Applied to materials A, B, and C, this criterion allows the use of the rigid frame model above 13 881 Hz for material A, 3041 Hz for material B, and 7535 Hz for material C. As stated previously, the parameter FSI allows the use of the limp model in the whole frequency range for both fibrous materials A and B, and recommends the use of the Biot model in the vicinity of the frame resonance for the material C. Hence, in this case, the frame can be excited by the motion of the air (predictable with the criterion f_{Biot}) but, since the stiffness has no influence (predictable with FSI), the "equivalent fluid" limp model can nevertheless be used.

Pilon et al.[16] proposed a criterion to study the influence of the mounting conditions on the measurement of the sound absorption coefficient in a duct. This criterion, called FAE and defined as FAE$=\sigma E/\rho_1^2$, gives informations about the influence of the frame around its frequency resonance f_r. For material C, the FAE is equal to 5.7 MW/kg and greatly superior to the critical value (1 MW/kg), which indicates that the acoustic behavior of the foam is very sensitive to the frame vibration in the vicinity of f_r. The prediction is in good agreement with the FSI criterion.

VII. CONCLUSION

In this paper, a criterion identifying which porous materials can be modeled with the one-wave limp model instead of the poroelastic Biot model has been investigated. The identification process relies on a new parameter derived from the properties of the material: the frame stiffness influence. This parameter, based on the compressional wave numbers, expresses the influence of the frameborne wave on the fluid phase displacement. The method consists in evaluating the FSI for a given material and comparing it with critical values determined for two characteristic configurations: absorption of a porous layer bonded on a rigid backing and radiation of a vibrating plate covered by a porous layer. It is shown that the frequency bands predicted to be simulated with the limp or the Biot model using this method are in close agreement

with one-dimensional simulations. Even if the application cases are one dimensional, they involve both Biot compressional waves. It seems realistic to apply this criterion to three-dimensional cases where the shear wave is not mainly excited, which is the case for a wide variety of industrial applications. Compared with other criteria, the FSI criterion provides a more accurate information in terms of frequency and confirms that the limp model is less restrictive than the rigid frame one.

ACKNOWLEDGMENTS

[1] M. A. Biot, "The theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low frequency range. II. Higher frequency range," J. Acoust. Soc. Am. **28**, 168–191 (1956).

[2] J. F. Allard, *Propagation of Sound in Porous Media: Modelling Sound Absorbing Materials* (Elsevier Applied Science, London, 1993).

[3] R. Panneton and N. Atalla, "An efficient scheme for solving the three-dimensional poroelasticity problem in acoustics," J. Acoust. Soc. Am. **101**, 3287–3298 (1998).

[4] N. E. Hörlin, M. Nordström, and P. Göransson, "A 3-D hierarchical FE formulation of Biots equations for elastoacoustic modeling of porous media," J. Sound Vib. **254**, 633–652 (2001).

[5] N. Atalla, R. Panneton, and P. Debergue, "A mixed displacement-pressure formulation for poroelastic materials," J. Acoust. Soc. Am. **104**, 1444–1452 (1998).

[6] S. Rigobert, N. Atalla, and F. Sgard, "Investigation of the convergence of the mixed displacement pressure formulation for three-dimensional poroelastic materials using hierarchical elements," J. Acoust. Soc. Am. **114**, 2607–2617 (2003).

[7] L. L. Beranek, "Acoustical properties of homogeneous, isotropic rigid tiles and flexible blankets," J. Acoust. Soc. Am. **19**, 556–568 (1947).

[8] K. U. Ingard, *Notes on Sound Absorption Technology* (Noise Control Foundation, New York, 1994).

[9] O. Dazel, B. Brouard, C. Depollier, and S. Griffiths, "An alternative Biots displacement formulation for porous materials," J. Acoust. Soc. Am. **121**, 3509–3516 (2007).

[10] C. Zwikker and C. W. Kosten, *Sound Absorption Materials* (Elsevier Applied Science, New York, 1949).

[11] P. Göransson, "A weighted residual formulation of the acoustic wave propagation through a flexible porous material and a comparison with a limp material model," J. Sound Vib. **182**, 479–494 (1995).

[12] P. Göransson, "Acoustic finite element formulation of a flexible porous material: A correction for inertial effects," J. Sound Vib. **185**, 559–580 (1995).

[13] D. L. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–402 (1987).

[14] Y. Champoux and J. F. Allard, "Dynamic tortuosity and bulk modulus in air-saturated porous media," J. Appl. Phys. **70**, 1975–1979 (1991).

[15] A. F. Seybert and D. F. Ross, "Experimental determination of acoustic properties using a two-microphone random-excitation technique," J. Acoust. Soc. Am. **61**, 1362–1370 (1977).

[16] D. Pilon, R. Panneton, and F. Sgard, "Behavioral criterion quantifying the edge-constrained effects on foams in the standing wave tube," J. Acoust. Soc. Am. **114**, 1980–1987 (2003).

[17] C. E. Wallace, "Radiation resistance of a rectangular panel," J. Acoust. Soc. Am. **51**, 946–952 (1972).

[18] O. Doutres, N. Dauchez, and J. M. Génevaux, "Porous layer impedance applied to a moving wall: Application to the radiation of a covered piston," J. Acoust. Soc. Am. **121**, 206–213 (2007).

[19] B. Brouard, D. Lafarge, and J. F. Allard, "A general method of modeling sound propagation in layered media," J. Sound Vib. **183**, 129–142 (1995).

[20] N. Dauchez, M. Etchessahar, and S. Sahraoui, "On measurement of mechanical properties of sound absorbing materials," Second Biot Conference on Poromechanics, Grenoble, 2002.

[21] M. Etchessahar, S. Sahraoui, L. Benyahia, and J. F. Tassin, "Frequency dependence of elastic properties of acoustic foams," J. Acoust. Soc. Am. **117**, 1114–1121 (2005).

[22] J. F. Allard, C. Depollier, P. Guignouard, and P. Rebillard, "Effects of a resonance of the frame on the surface impedance of glass wool of high density and stiffness," J. Acoust. Soc. Am. **89**, 999–1001 (1991).

Guided wave modes in porous cylinders: Theory

C. J. Wisse
Department of Geotechnology, Delft University of Technology, P.O. Box 5028, 2600 GA, Delft,
The Netherlands and Department of Applied Physics, Eindhoven University of Technology, P.O. Box 513,
5600 MB, Eindhoven, The Netherlands

D. M. J. Smeulders
Department of Geotechnology, Delft University of Technology, P.O. Box 5028, 2600 GA, Delft,
The Netherlands

G. Chao
Department of Applied Physics, Eindhoven University of Technology, P.O. Box 513, 5600 MB, Eindhoven,
The Netherlands

M. E. H. van Dongen
Department of Applied Physics, Eindhoven University of Technology, P.O. Box 513, 5600 MB, Eindhoven,
The Netherlands

The classical theory of wave propagation in elastic cylinders is extended to poro-elastic mandrel modes. The classical theory predicts the existence of undamped L modes and damped C, I, and Z modes. These waves also appear in poro-elastic mandrels, but all of them become damped because of viscous effects. The presence of the Biot slow bulk wave in the poro-elastic material is responsible for the generation of additional mandrel modes. One of them was already discussed by Feng and Johnson, and the others can be grouped together as so-called D modes. The damping of these D modes is at least as high as the damping of the free-field slow wave. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2767418]

I. INTRODUCTION

For the interpretation of acoustic borehole logs a thorough understanding of the effects of rock properties on acoustic wave propagation is required. Examples of these properties are the porosity, permeability, lithology, and hydrocarbon saturation. In conventional acoustic borehole logging the first arrival of the acoustic wave train can be used to determine the porosity. The full wave train is applied to determine mechanical properties, while one of the late arrivals, the Stoneley wave, is used to obtain permeability data.[1,2] The properties of the different arrivals in the time domain can be studied separately.

Rosenbaum[3] used Biot's theory[4,5] to describe the wave modes in a borehole. Schmitt *et al.*,[6] Liu,[7] and Winkler *et al.*[8] demonstrated the permeability dependence of the Stoneley wave. It was argued that the permeability dependence of the Stoneley wave could be explained by its relation to the slow Biot wave.

The present study is concerned with wave modes in a porous cylinder. Special attention is given to the influence of the slow Biot wave on the wave modes. The configuration is given in Fig. 1, representing a porous mandrel inside a steel tube. Between the porous cylinder and the inner wall of the tube a water-filled annulus exists. Experimental results for this configuration were discussed in a previous paper.[9] In these experiments, a series of pressure waveforms is recorded at different positions along the cylinder. A FFT-Prony method is implemented to transform the data from the time-space domain to the frequency-wavenumber domain. In this way, frequency-dependent velocities and attenuation of individual mandrel modes can be studied separately. The fundamental mode is the Stoneley wave propagating along the interface between the porous medium and the liquid. However, also higher-order modes were observed in some of our previous experiments.[10] This observation motivates our systematic study of these higher-order modes, which may appear in poro-elastic mandrels.

First, the wave modes in nonporous elastic mandrels are revisited. Wave propagation in elastic cylinders is described by the Pochhammer equation, which has been investigated by many researchers.[11,12] In essence, this equation describes the propagation of wave modes through an elastic cylinder (mandrel) with free-surface boundary conditions. A finite number of undamped modes occur, as well as an infinite number of strongly damped modes.[13,14]

Gardner[15] was the first to study wave modes in porous cylinders. In that work the long-wavelength approximation was combined with the low-frequency limit of Biot's theory. Two types of wave modes were predicted, one related to the free-field fast wave and one related to the free-field slow wave. Berryman[16] used the full Biot theory to study the wave modes in a poro-elastic cylinder. When the pores of the cylinder have an open connection with the outside world (open-pore condition), three wave modes were found that are similar to the wave modes in an elastic cylinder. No waves associated with the free-field slow wave were found. When the pores are sealed from the outside world (closed-pore con-
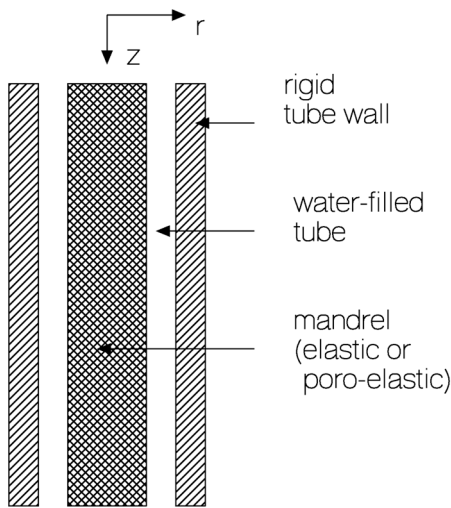
FIG. 1. (Color online) Schematic diagram for computing the characteristics of axially symmetric modes is a water-filled tube with a mandrel that may be either elastic or poro-elastic. The mandrel radius is a, the inner tube radius is b. The configuration is presumed to be of infinite length.

dition), a wave type was found that had several properties of the free-field slow wave. Its velocity, however, was lower than that of the free-field slow wave. It was suggested that this wave type is similar to the true surface wave discussed by Feng and Johnson.[17] Hsu et al.[18] investigated wave modes in a (porous) mandrel in a cylindrical borehole surrounded by an elastic formation. Their configuration is similar to ours. A wave mode that could only be explained by the presence of the slow Biot wave was found. Further work on porous cylinders has been carried out by Johnson and Kostek,[19] who provided a comparison between the approximation of Gardner and the full Biot theory for the first mode type. Previous papers have in common that usually one or two waves are discussed that are of importance for practical purposes. In this paper, the full set of wave modes that exist in poro-elastic cylinders is discussed. In this respect, it is an extension of the work by Zemanek[14] to the poro-elastic case.

The approach of this paper is as follows. First, the results for the undamped and damped modes in a nonporous elastic cylinder are briefly reviewed. Next, the computational results are given for a water-saturated mandrel made out of natural Bentheimer sandstone, which is a typical example of the class of consolidated sandstones.

II. ELASTIC CYLINDERS

A detailed analysis of the wave modes in an elastic cylinder is given by Onoe et al.[13] and Zemanek.[14] Following the approach of Miklowitz[12] for the axial symmetric case, the following decomposition for the displacement vector $\mathbf{u}_s$ is used:

$$\mathbf{u}_s = \nabla \varphi_s + \nabla \times \nabla \times (\eta_{sh}\mathbf{e}_z), \tag{1}$$

where the potentials φ_s and η_{sh} satisfy the wave equation with respect to the compressional and shear wave velocity, respectively. The unit vector in the z-direction is $\mathbf{e}_z$. For cylindrical coordinates with z as the axial coordinate and r as the radial coordinate it can be derived that[12]
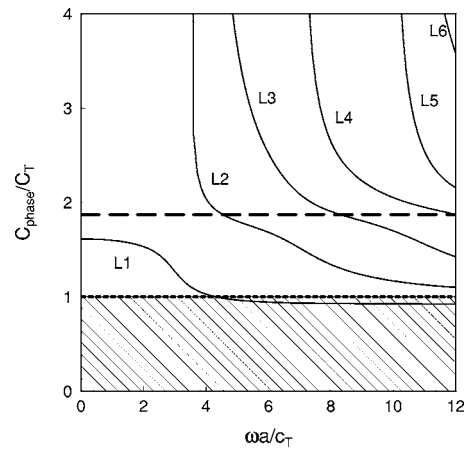
FIG. 2. Undamped modes: Frequency dependence of the phase velocity for an elastic cylinder with $\nu=0.3$. Dashed line: $\omega=C_L\mathrm{Re}(k)$. Shaded region: $\omega/C_T < \mathrm{Re}(k)$. For reference, also the free-field shear (dotted line) and compressional (dashed line) waves are shown.

$$\varphi = A_{c1}J_0(k_{rc1}r)e^{i(\omega t - kz)} \tag{2}$$

and

$$\eta_{sh} = C_{sh}J_0(k_{rsh}r)e^{i(\omega t - kz)}. \tag{3}$$

The radial wavenumbers k_{rc1} and k_{rsh} are defined as

$$k_{rc1} = \sqrt{\frac{\omega^2}{C_L^2} - k^2}, \quad k_{rsh} = \sqrt{\frac{\omega^2}{c_T^2} - k^2}. \tag{4}$$

In these expressions c_L and c_T denote the compressional and shear velocity, respectively. The free-surface condition requires that all stresses at $r=a$ vanish:

$$\tau_{rr} = 0, \quad \tau_{rz} = 0. \tag{5}$$

For the amplitude coefficients the following system of equations is obtained:

$$\underline{\underline{\mathbf{M}}}(A_{c1}, C_{sh})^T = \underline{0}, \tag{6}$$

where the elements of $\underline{\mathbf{M}}$ can be derived from the boundary conditions (see Ref. 12, pp. 220–221). The requirement that the determinant of coefficients must vanish yields the frequency equation: $F(\omega,k)=0$. This equation is the well-known Pochhammer frequency equation. This equation is solved using a Newton-Raphson algorithm. At a given frequency ω, a finite number of undamped modes and an infinite number of damped modes exist.[14]

The results of the calculations for the undamped modes are shown in Fig. 2. On the horizontal axis the dimensionless frequency is shown, while on the vertical axis the dimensionless phase velocity is given, where a is the radius of the cylinder. Pochhammer's equation is determined uniquely by the variables $\omega a/c_T$, ka, and Poisson's ratio ν.[11] The results are given for $\nu=0.3$. For comparison, the free-field velocities of the compressional and the shear waves are also given. The solutions in the dashed region correspond to surface waves. In Fig. 2, one wave mode, which is propagative in the entire frequency range, is observed. This mode is labeled L1. In the low-frequency limit the phase velocity of the L1 mode is called the bar velocity. In the high-frequency limit the phase velocity is equal to the velocity of the Rayleigh wave on a
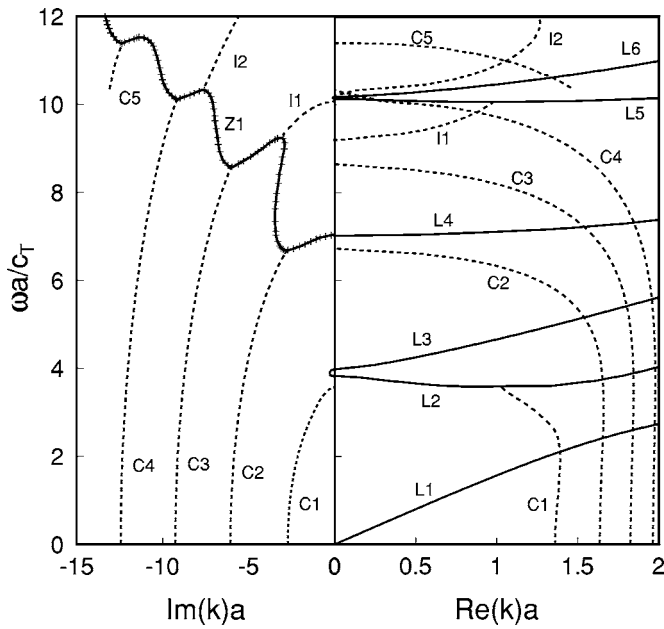
FIG. 3. Modes of an elastic cylinder in the complex plane. The solid lines correspond to the L modes of Fig. 2(a). The pairs of dashed lines correspond to the imaginary and real parts of k. The computations were performed for an elastic cylinder with $\nu = 0.3$.

TABLE I. Parameter values.

Density of water ρ_w (kg/m^3)	998
Dynamic viscosity of water η (Pa s)	10^{-3}
Water bulk modulus K_f (GPa)	2.2
Density of the solid ρ_s (10^3 kg/m^3)	2.62
Porosity (%)	0.22
Permeability k_0 (10^{-12} m^2)	2.73
Tortuosity	2.9
Constrained modulus K_p (GPa)	17
Shear modulus G (GPa)	8.0

flat interface. At these high frequencies, the wavelength becomes large enough so that the curvature of the surface is negligible with respect to the wavelength. The higher-order modes $L2, \ldots, L6$ are characterized by the high-frequency limit of the phase velocity, which is the free-field shear velocity. At lower frequencies they show cutoff behavior, i.e., below the cutoff frequencies these modes are nonpropagating.

For the same cylinder, also an infinite number of damped modes exist, for which $\mathrm{Re}(k)a$ and $\mathrm{Im}(k)a$ are given in Fig. 3 as a function of the frequency $\omega a/c_T$. For comparison, also the undamped wave modes are drawn (solid lines). Modes that appear only on the right-hand side of Fig. 3 are propagatory and undamped. Modes that appear only on the left-hand side of Fig. 3 [the $\mathrm{Re}(k)a - \omega a/c_T$ part] are nonpropagative and damped, whereas modes that appear in both plots are propagating damped waves. The latter are indicated by pairs of dashed lines: one in the $\mathrm{Re}(k)$ plane, and a corresponding one in the $\mathrm{Im}(k)$ plane. Labels C and I are used here. The C modes exist at zero frequency and for higher frequencies they catch up to the Z1 mode, which is a nonpropagative mode. The I modes originate from the Z1 mode and eventually catch up to the nonpropagative Z2 mode (not shown in Fig. 3). The C, I, and Z modes play an important role near the ends of cylinders with finite length,[14] but the discussion of end aspects is beyond the scope of the present paper.

III. POROUS CYLINDER IN THE SHOCK TUBE

The configuration of the liquid-loaded water-saturated porous cylinder is given in Fig. 1. The cylinder is considered infinitely long. The radius of the porous cylinder is denoted by a, while the inner radius of the surrounding tube is denoted by b. Biot's theory is used to describe wave propaga-

tion in the porous cylinder. A potential formulation based on Biot's theory is used to solve the wave equation in cylindrical coordinates in the frequency domain. This formulation takes into account the two free-field compressional waves and the free-field shear wave in the porous cylinder as well as the compressional wave that propagates in the fluid. The boundary conditions lead to a linear system for the amplitudes of the waves. In order to find the wave modes the determinant of this system must be equal to zero. This condition leads to a secular equation for the wavenumber of the guided wave. The secular equation is solved using a zero search routine based on Newton-Raphson's iteration method in the complex plane. At $r=a$ the open-pore boundary conditions as given by Deresiewicz and Skalak[20] and Rosenbaum[3] are applied, while at $r=b$ the radial displacement of the water is assumed to be zero. In our experiments, this is a realistic assumption indeed, as the walls of the shock tube are made of 25-mm-thick stainless steel. Hence, the following boundary conditions apply:

(1) at $r=a$: $\sigma_{rr}=0$,
(2) at $r=a$: $\sigma_{rz}=0$,
(3) at $r=a$: $(1-\phi)u_{sr}+\phi u_{fr}=u_{rw}$,
(4) at $r=a$: $p=p_w$,
(5) at $r=b$: $u_{r2}=0$.

Using the expressions for the displacements and stresses as given in Appendix A, the five unknown amplitudes can be written as

$$\underline{\underline{M}}(A_{c1}, \quad A_{c2}, \quad C_{sh}, \quad B_w, \quad A_w)^T = \underline{0}. \tag{7}$$

The elements of matrix $\underline{\underline{M}}$ are given in Appendix B.

IV. COMPUTATIONAL RESULTS FOR PORO-ELASTIC CYLINDERS

The computations are performed for a water-saturated Bentheim sandstone whose properties are given in Table I. This is a fast formation, i.e., the shear velocity is higher than the water velocity. The computations were performed with a diameter of the cylinder of 70 mm, while the inner diameter of the tube is 77 mm, i.e., $a=0.909b$ (see Fig. 1).

Some of the modes that were found show similarities with the L, C, and I modes of the nonporous elastic cylinders examined in Sec. III, and consequently these modes are labelled L, C, and I here as well. Furthermore, a surface mode is found, which will be denoted by S. Due to the presence of the free-field Biot slow wave in the porous cylinder, an extra class of modes occurs, which is denoted by D. Because of

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Wisse *et al.*: Poro-elastic mandrel modes   2051
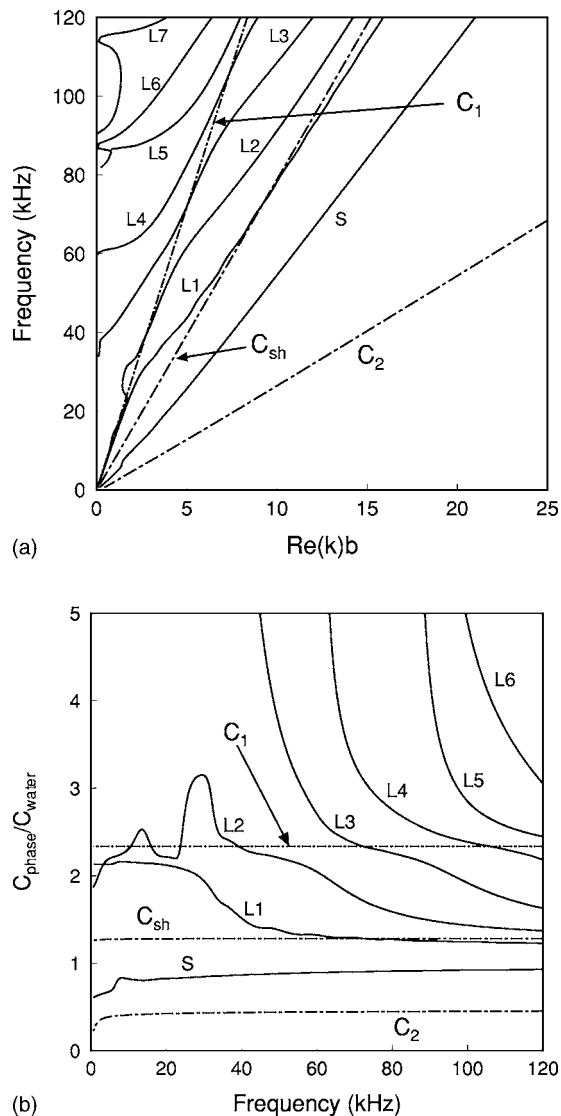
FIG. 4. Frequency dependence of Re(k) (a) and the phase velocity (b) for the L and S modes. The wavenumber is normalized by the inner radius of the shock tube b, the phase velocity by the sound velocity of water. The free-field solutions for the fast (C_1), the slow (C_2), and the shear wave (C_{sh}) are also given. The computations are performed for a Bentheimer sandstone cylinder with a diameter of 70.0 mm. Note that the full computational results are presented in three figures. L and S modes are presented here, the C and I modes are shown in Fig. 7, and the D modes are shown in Fig. 8.

the complexity of the complete mode pattern the results are presented in separate figures. First, the L and S modes will be discussed, which are characterized by a relatively low damping factor. These results are shown in Fig. 4, while the results for the C, I, and D modes are shown in Figs. 7 and 8.

A. L and S modes

The data for the L and S modes are presented in the same way as for the nonporous elastic case, but, for practical purposes,[9] the frequency instead of the dimensionless parameter $\omega a / c_T$ is used. In Fig. 4(a) the frequency versus the real part of k is plotted, while in Fig. 4(b) the frequency dependence of the phase velocities is given.

Note that Re(k) is normalized by the inner radius of the shock tube b, while the phase velocities are normalized by

the sound velocity of water. For comparison, also the free-field solutions for the fast (C_1), slow (C_2), and shear wave (C_{sh}) are given.

The L1 and L2 modes have a finite phase velocity over the entire frequency range, while the higher-order L modes are cut off below a certain frequency. The L modes in Fig. 4 show strong similarities with the results of Fig. 2, where one mode that is propagative over the entire frequency range is observed, while the other modes were cut off below a critical frequency. There is, however, an important difference with respect to the case of nonporous elastic cylinders. The L1 mode for the nonporous elastic cylinder is undamped over the entire frequency range, and the higher-order L modes are undamped above the cutoff frequency. For the porous material, however, the free-field fast, slow, and shear waves are damped, and, consequently, all the L modes are damped over the entire frequency range. The damping coefficients of the L1 mode over the entire frequency range, and of the higher-order L modes above the cutoff frequency, are low, i.e., $|\text{Im}(k)b| \ll 1$ (not plotted). The high-frequency velocity of the L1 mode equals the sound speed of water and the damping tends to zero. This wave is therefore associated with wave motion in the annulus. In the low-frequency limit, the phase velocity is somewhat lower than the free-field velocity of the fast wave. For the purely elastic case, this limiting speed was the so-called bar velocity $\sqrt{E/\rho}$. Further analysis showed that for this mode radial and axial stress components are of equal importance at high frequencies.[21] In the low-frequency limit the one-dimensional stress situation is obtained, i.e., the axial stress is dominant over the radial and shear stresses, and pore pressures are of the same order as the pressure in the annulus.

In the high-frequency limit, the L2 phase velocity reaches the phase velocity of the free-field shear velocity. At low frequencies the phase velocity remains finite, which is different from the behavior of the L2 mode for the nonporous elastic case. The phase velocity of the latter mode becomes infinite at the cutoff frequency. Note also that the damping of the L2 mode for the porous cylinder is relatively large at low frequencies ($|\text{Im}(k)b| > 1$) (see Fig. 7).

Now the S mode is further analyzed. This wave is generally conspicuously present in our experiments[9] and therefore merits further investigation. Figure 5 displays the frequency dependence of the phase velocity and the damping coefficient of the S mode in the 1–1000-kHz frequency range. For comparison the pseudo-Stoneley wave (denoted PS$_{41}$ in the figure) for a plane interface is also shown.[21] From Fig. 5 it is observed that the S mode is equivalent to the plane-interface pseudo-Stoneley wave in the high-frequency range, where the wavelengths involved are much shorter than the radii of the shock tube and the porous cylinder. Both the phase velocities and the damping coefficients are nearly identical. At lower frequencies, the S mode is influenced by the curvatures of the porous cylinder and of the wall of the shock tube, as the wavelength becomes of the order of magnitude of the radii of the shock tube and the porous cylinder.

Figure 6 shows the radial distributions of the stresses and pressures for the S mode at 120 kHz. The stresses and pressures are normalized by the pore pressures at the radial
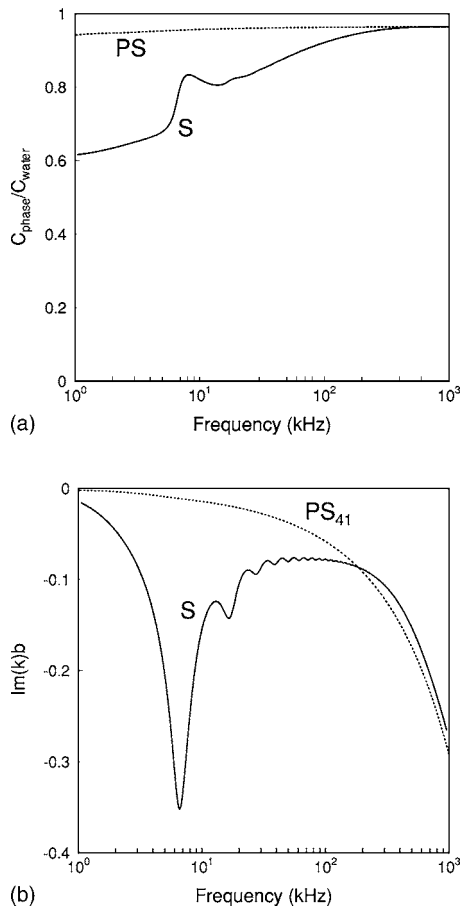
FIG. 5. Frequency dependence of the phase velocity (a) and damping (b) for the S mode in the 1–1000-kHz frequency range. The results for the pseudo-Stoneley wave for a flat interface are also shown. The computations are performed for a Bentheimer sandstone cylinder with a diameter of 70.0 mm. Parameter values are given in Table I.

FIG. 6. Radial distributions of the stresses (a) and pressures (b) for the S mode at 120 kHz. Both the stresses and pressures were normalized by $p_p|_{r=a}$. The calculations were performed for a Bentheim cylinder with a diameter of 70.0 mm ($a=0.909b$). The pores at the boundary were open.

surface of the porous cylinder $r=a$. We notice in Fig. 6 that the typical behavior of the surface wave is present at 120 kHz: a pressure peak at the wall is observed, followed by an exponential-like decay. The pressure oscillations for r/b values below 0.4 are generated by the free-field slow wave. We found that as the frequency is lowered, the pressure peak at the wall becomes broader, and the surface-wave character disappears. It can be argued that the S mode is associated with a surface wave at high frequencies, while it corresponds to a bulk wave type at low frequencies. Note that this is similar to the corresponding mode in a borehole, which is associated with the so-called "water-hammer" at low frequencies and with the pseudo-Stoneley wave at high frequencies. Furthermore, it can be remarked that over the entire frequency range, the annulus pressures are of the same order as the pore pressures in the porous cylinder.

Summarizing, from the analysis of the S, L1, and L2 modes, it can be stated that for the S and L1 modes the motion of the water in the annulus plays an important role. The annulus pressures for these modes are of the same order of magnitude as the stresses and pressures in the porous cylinder. The L2 mode, however, is dominated by the motion of the solid matrix. The same behavior can be expected for the higher-order wave modes. It is worthwhile to mention that in

some of our previous experiments we observed the excitation of L modes in a borehole configuration as well.[10]

B. C and I modes

Our numerical computations also predict a number of strongly damped modes, which are similar to the damped C and I modes that we observed in computing the nonporous elastic case (see Fig. 3). Consequently, these modes for the poro-elastic cylinder are labelled C and I as well. In Fig. 7, the results are presented in the same way as for the nonporous elastic case. The C modes are indicated by dashed lines, and the I modes are indicated by solid lines. For comparison, some of the L modes that were already presented in Fig. 4 are plotted again. The behavior of $\text{Re}(k)$ of the C modes for the porous cylinder is somewhat different from the C modes for the elastic case, which have a finite value of $\text{Re}(k)$ at low frequencies. For the C modes of the porous cylinder, $\text{Re}(k)$ is zero in the low-frequency limit. The C modes are connected to the I modes via modes in the third quadrant of the k plane. Those modes have infinite amplitudes for $z \rightarrow \infty$, and are not shown in Fig. 7.
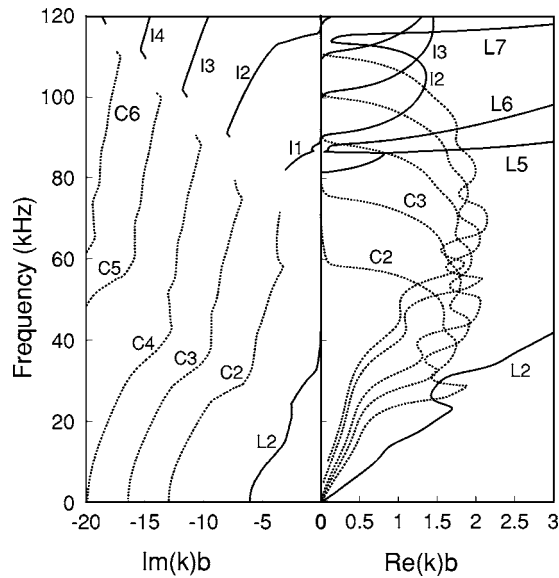
FIG. 7. Frequency dependence of Im(k) and Re(k) for the C and I modes. As a reference some L modes are given. The wavenumber is normalized by the inner radius of the shock tube b. The computations were performed for a Bentheim cylinder with a diameter of 70.0 mm. The pores at the boundary were open. Note that the full computational results are presented in three figures. The L and S modes are shown in Fig. 4, while the D modes are shown in Fig. 8.

FIG. 8. Frequency dependence of Im(k) and Re(k) for the D modes. The wavenumber is normalized by the inner radius of the shock tube b. The computations were performed for a Bentheimer sandstone cylinder with a diameter of 70.0 mm. The pores at the boundary were open. The free-field solution C_2 for the slow wave is indicated by the dashed line. Note that the full computational results are presented in three figures. The L and S modes are shown in Fig. 4, while the C and I modes are shown in Fig. 7.

C. D modes

Up to now it was possible to relate our computational results to the nonporous elastic cylinder. For the poro-elastic cylinder in the shock tube, however, an extra mode type has been found, closely related to the existence of the free-field slow wave. The occurrence of this wave mode family for open-pore boundary conditions is in agreement with the predictions by Gardner,[15] Liu,[7] and Hsu *et al.*[18] In the following, the slow wave mode is denoted by D. In fact, this mode is essentially a slow wave with some coupling to the other mode types. The results for the D modes are shown in Fig. 8, where the frequency dependence of the real and imaginary parts of k is shown. In this figure the dashed lines indicate the free-field solution for the slow wave. From these results it is clearly observed that the D modes have the free-field solution of the slow wave as their high-frequency limit. The damping of these modes is at least as high as the damping of the slow wave and of the same order of magnitude as the damping of the C modes.

The pore pressures and the stress components are of the same order of magnitude, while the pressures in the center of the cylinder are much larger than the gap pressure. This means that the wave motion in the gap is of less importance for the D modes.[21]

V. CONCLUSIONS

Porous mandrels are investigated. The so-called L modes are equivalent to the ones in a nonporous elastic cylinder. These modes are characterized by cutoff frequencies and their phase velocities in the high-frequency limit, which are equal to the shear velocity. For the porous cylinder, an equivalent yet strongly damped set of modes is found. Fur-

thermore, a surface mode S occurs at the interface between a porous cylinder and a relatively large water-filled gap. At high frequencies this wave corresponds to the pseudo-Stoneley wave. Extra bulk wave modes, the D modes, occur due to the presence of the Biot slow wave. The damping of the higher-order D modes is at least as high as the damping of the Biot slow wave itself.

APPENDIX A: BIOT THEORY IN POROUS CYLINDERS

The potential φ_s in Eq. (1) consists of two parts, one for the fast Biot wave and one for the slow Biot wave:[16,22]

$$\varphi_s = \varphi_{sc1} + \varphi_{sc2}. \tag{A1}$$

The fast wave potential satisfies the wave equation with respect to the complex-valued velocity of the fast wave c_1 and the slow wave potential satisfies the wave equation with respect to the complex-valued velocity of the slow wave c_2. The shear potential η_{sh} satisfies the wave equation with respect to the complex-valued velocity of the shear wave c_{sh}. For φ_{sc1}, φ_{sc2}, and η_{sh} it is possible to derive that

$$\varphi_{sc1} = A_{c1}J_0(k_{rc1}r)e^{i(\omega t - kz)}, \tag{A2}$$

$$\varphi_{sc2} = A_{c2}J_0(k_{rc2}r)e^{i(\omega t - kz)}, \tag{A3}$$

and

$$\eta_{sh} = C_{sh}J_0(k_{rsh}r)e^{i(\omega t - kz)}. \tag{A4}$$

The radial wavenumbers are defined by

$$k_{rc1} = \sqrt{\frac{\omega^2}{c_1^2} - k^2},$$

$$k_{rc2} = \sqrt{\frac{\omega^2}{c_2^2} - k^2}, \tag{A5}$$

$$k_{rsh} = \sqrt{\frac{\omega^2}{c_{sh}^2} - k^2},$$

with $\mathrm{Im}(k_{rc1}) \leq 0$, $\mathrm{Im}(k_{rc2}) \leq 0$, and $\mathrm{Im}(k_{rsh}) \leq 0$. The potentials for the fluid are related to the potentials of the solid by the following expressions:

$$\varphi_{fc1} = \beta_{c1}\varphi_{sc1}, \quad \varphi_{fc2} = \beta_{c2}\varphi_{sc2}, \quad \eta_{fsh} = \beta_{sh}\eta_{sh}. \tag{A6}$$

The coefficients β_{c1}, β_{c2}, and β_{sh} are frequency-dependent and can be derived from the equations of motion (see Ref. 22, p. 131).

The displacements of the solid and the fluid can be derived from the potentials by the following relations:

$$u_{sr} = \frac{\partial \varphi_{sc1}}{\partial r} + \frac{\partial \varphi_{sc2}}{\partial r} + \frac{\partial^2 \eta_{sh}}{\partial z \, \partial r} \tag{A7}$$

and

$$u_{fr} = \beta_{c1}\frac{\partial \varphi_{sc1}}{\partial r} + \beta_{c2}\frac{\partial \varphi_{sc2}}{\partial r} + \beta_{sh}\frac{\partial^2 \eta_{sh}}{\partial z \, \partial r}. \tag{A8}$$

For the isotropic case the stress-strain relations are given by Biot.[23] Using cylindrical coordinates the radial stress component acting on the solid part can be written as

$$\tau_{rr} = (P - 2G)\nabla^2\varphi_s + Q\nabla^2\varphi_f + 2G\left[\frac{\partial^2 \varphi_s}{\partial r^2} - \frac{\partial^3 \eta_{sh}}{\partial z \, \partial r^2}\right], \tag{A9}$$

where the elasticity coefficients P, Q, R, and G have been introduced. The relation between these coefficients and measurable quantities are given by Allard.[22] For the shear stress τ_{rz} it follows that

$$\tau_{zr} = G\left\{2\frac{\partial^2 \varphi_s}{\partial z \, \partial r} + \frac{\partial^3 \eta_{sh}}{\partial z^2 \, \partial r} - \frac{\partial}{\partial r}\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \eta_{sh}}{\partial r}\right)\right]\right\}. \tag{A10}$$

The stress component acting on the fluid part can be written as

$$\tau = Q\nabla^2\varphi_s + R\nabla^2\varphi_f. \tag{A11}$$

The radial stress acting on the solid part can be related to the radial component of the so-called intergranular stress σ_{rr} and the pore pressure p_p by

$$\tau_{rr} = -\sigma_{rr} - (1-\phi)p_p, \tag{A12}$$

where ϕ is the porosity. For σ_{rr} it holds that

$$\sigma_{rr} = -\Lambda_{c1}\nabla^2\varphi_{sc1} - \Lambda_{c2}\nabla^2\varphi_{sc2} - 2G\left[\frac{\partial^2 \varphi_s}{\partial r^2} - \frac{\partial^3 \eta_{sh}}{\partial z \, \partial r^2}\right], \tag{A13}$$

where

$$\Lambda_{c1} = P - 2G - \frac{(1-\phi)Q}{\phi} + Q\beta_{c1} - \frac{(1-\phi)R\beta_{c1}}{\phi} \tag{A14}$$

and

$$\Lambda_{c2} = P - 2G - \frac{(1-\phi)Q}{\phi} + Q\beta_{c2} - \frac{(1-\phi)R\beta_{c2}}{\phi}. \tag{A15}$$

For intergranular shear stress it is clear that $\sigma_{rz} = \tau_{rz}$. The pore pressure can be written as $p_p = -\tau/\phi$.

The wave motion in the water-filled gap is described by the wave equation of water. Hence, it is possible to derive the potential of the water-filled gap:

$$\varphi_w = [B_2 Y_0(k_{rw}r) + A_2 J_0(k_{rw}r)]e^{i(\omega t - kz)}, \tag{A16}$$

where the Bessel function of the second kind Y_0 is included and

$$k_{rw} = \sqrt{\frac{\omega^2}{c_w^2} - k^2} \tag{A17}$$

with $\mathrm{Im}(k_{rw}) \leq 0$. The pressure in the liquid can be written as

$$p_w = \rho_w \omega^2 \varphi_w \tag{A18}$$

and the radial water displacement as

$$u_{wr} = \frac{\partial \varphi_w}{\partial r} = -k_{rw}B_w Y_1(k_{rw}r) - k_{rw}A_w J_1(k_{rw}r), \tag{A19}$$

where the $e^{i(\omega t - kz)}$ dependence has been omitted.

APPENDIX B: MATRIX ELEMENTS FOR THE POROUS CYLINDER

The boundary conditions for the porous cylinder in the shock tube give the following system of equations for the amplitudes of the potentials:

$$\underline{\mathbf{M}}(A_{c1}, \quad A_{c2}, \quad C_{sh}, \quad B_w, \quad A_w)^T = 0. \tag{B1}$$

The elements of matrix $\underline{\mathbf{M}}$ are given by

$$M(1,1) = \Lambda_{c1}\left(\frac{\omega^2}{c_1^2}\right)J_0(k_{rc1}a) + 2G\left[k_{rc1}\left(k_{rc1}J_0(k_{rc1}a)\right. \right.$$
$$\left.\left. -\frac{1}{a}J_1(k_{rc1}a)\right)\right],$$

$$M(1,2) = \Lambda_{c2}\left(\frac{\omega^2}{c_2^2}\right)J_0(k_{rc2}a) + 2G\left[k_{rc2}\left(k_{rc2}J_0(k_{rc2}a)\right. \right.$$
$$\left.\left. -\frac{1}{a}J_1(k_{rc2}a)\right)\right],$$

$$M(1,3) = -2Gikk_{rsh}\left(k_{rsh}J_0(k_{rsh}a) - \frac{1}{a}J_1(k_{rsh}a)\right),$$

$$M(1,4) = 0,$$

$$M(1,5) = 0,$$

$$M(2,1) = -2ikk_{rc1}J_1(k_{rc1}a),$$

$$M(2,2) = -2ikk_{rc2}J_1(k_{rc2}a),$$

$$M(2,3) = -k_{rsh}(k^2 - k_{rsh}^2)J_1(k_{rsh}a),$$

$$M(2,4) = 0,$$

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Wisse *et al.*: Poro-elastic mandrel modes   2055

$M(2,5) = 0,$

$M(3,1) = ((\phi - 1) - \beta_{c1}\phi)k_{rc1}J_1(k_{rc1}a),$

$M(3,2) = ((\phi - 1) - \beta_{c2}\phi)k_{rc2}J_1(k_{rc2}a),$

$M(3,3) = ((1 - \phi) + \beta_{sh}\phi)ikk_{rsh}J_1(k_{rsh}a),$

$M(3,4) = k_{rw}Y_1(k_{rw}a),$

$M(3,5) = k_{rw}J_1(k_{rw}a),$

$M(4,1) = \frac{1}{\phi}\left(\frac{\omega^2}{c_1^2}\right)J_0(k_{rc1}a)(Q + R\beta_{c1}),$

$M(4,2) = \frac{1}{\phi}\left(\frac{\omega^2}{c_2^2}\right)J_0(k_{rc2}a)(Q + R\beta_{c2}),$

$M(4,3) = 0,$

$M(4,4) = -\rho_w\omega^2 Y_0(k_{rw}a),$

$M(4,5) = -\rho_w\omega^2 J_0(k_{rw}a),$

$M(5,1) = 0,$

$M(5,2) = 0,$

$M(5,3) = 0,$

$M(5,4) = -k_{rw}Y_1(k_{rw}b),$

$M(5,5) = -k_{rw}J_1(k_{rw}b).$

The matrix elements only have combinations of $J_0(k_{rj}r)$, $k_{rj}^2 J_0(k_{rj}r)$, and $k_{rj}J_1(k_{rj}r)$, where $j = c1$, $c2$, sh. These combinations are even functions of the corresponding radial wavenumbers and therefore the choice of the Riemann sheet for the square root is arbitrary. The Neumann functions $Y_0(x)$ and $Y_1(x)$ are neither an odd nor an even function of x. However, $Y_0(-x)$ can be written as the sum of $Y_0(x)$ and $J_0(x)$.[24] Hence, the final result of the potential φ_w as given by Eq. (A16) is not influenced by the sign of k_{rw}, and the choice of the Riemann sheet for the radial wavenumber k_{rw} is arbitrary. The only relevant cut is the branch cut of the functions Y_0 and Y_1: $|\arg(k_{rw}r)| = \pi$. Passing this cut can be avoided by using $\mathrm{Im}(k_{rw}) \leq 0$.

[1] F. L. Paillet and C. H. Cheng. *Acoustic Waves in Boreholes—The Theory and Application of Acoustic Full-Waveform Logs* (CRC, Boca Raton, FL, 1991).

[2] X. M. Tang and C. H. Cheng, "Fast inversion of formation permeability from Stoneley wave logs using a simplified Biot-Rosenbaum model," Geophysics **61**, 639–645 (1996).

[3] J. H. Rosenbaum, "Synthetic microseismograms: Logging in porous formations," Geophysics **39**, 14–32 (1974).

[4] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).

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

[6] D. P. Schmitt, M. Bouchon, and G. Bonnet, "Full-wave synthetic acoustic logs in radially semi-infinite saturated porous media," Geophysics **53**, 807–823 (1988).

[7] H. L. Liu, "Borehole modes in a cylindrical fluid-saturated permeable medium," J. Acoust. Soc. Am. **84**, 424–431 (1988).

[8] K. W. Winkler, H.-S. Liu, and D. L. Johnson, "Permeability and borehole Stoneley waves: Comparison between experiment and theory," Geophysics **54**, 66–75 (1989).

[9] C. J. Wisse, D. M. J. Smeulders, M. E. H. van Dongen, and G. Chao, "Guided wave modes in porous cylinders: Experimental results," J. Acoust. Soc. Am. **112**(3), 890–895 (2002).

[10] G. Chao, D. M. J. Smeulders, and M. E. H. van Dongen, "Measurements of shock-induced guided and surface acoustic waves along boreholes in poroelastic materials," J. Appl. Phys. **99**(9), 094904 (2006).

[11] J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam, 1973).

[12] J. Miklowitz, *The Theory of Elastic Waves and Waveguides* (North-Holland, Amsterdam, 1978).

[13] M. Onoe, H. D. McNiven, and R. D. Mindlin, "Dispersion of axially symmetric waves in elastic rods," J. Appl. Mech. **29**, 729–734 (1962).

[14] J. Zemanek, "An experimental and theoretical investigation of elastic wave propagation in a cylinder," J. Acoust. Soc. Am. **51**, 265–283 (1971).

[15] G. H. F. Gardner, "Extensional waves in fluid-saturated porous cylinders," J. Acoust. Soc. Am. **34**, 36–40 (1962).

[16] J. G. Berryman, "Dispersion of extensional waves in fluid-saturated porous cylinders at ultrasonic frequencies," J. Acoust. Soc. Am. **74**, 1805–1812 (1983).

[17] S. Feng and D. L. Johnson, "High-frequency acoustic properties of a fluid/porous solid interface," J. Acoust. Soc. Am. **74**, 906–924 (1983).

[18] C.-J. Hsu, S. Kostek, and D. L. Johnson, "Tube waves and mandrel modes: Experiment and theory," J. Acoust. Soc. Am. **102**, 3277–3289 (1997).

[19] D. L. Johnson and S. Kostek, "A limitation of the Biot-Gardner theory of extensional waves in fluid-saturated porous cylinders," J. Acoust. Soc. Am. **97**, 741–744 (1995).

[20] H. Deresiewicz and R. Skalak, "On uniqueness in dynamic poroelasticity," Bull. Seismol. Soc. Am. **53**, 783–788 (1963).

[21] C. J. Wisse, "On frequency dependence of acoustic waves in porous cylinders," Ph.D. thesis, Delft University of Technology, 1999.

[22] J. F. Allard, *Propagation of Sound in Porous Media* (Elsevier Science, Amsterdam, 1993).

[23] M. A. Biot, "Theory of elasticity and consolidation for a porous anisotropic solid," J. Appl. Phys. **26**, 182–185 (1955).

[24] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).

Point of impact prediction in isotropic and anisotropic plates from the acoustic emission data

Tribikram Kundu[a)] and Samik Das[b)]

Department of Civil Engineering and Engineering Mechanics, University of Arizona, Tucson, Arizona 85721

Kumar V. Jata[c)]

Air Force Research Laboratory, MLLP, NDE Branch, Wright Patterson Air Force Base, Ohio 45433

It is shown in this paper that the conventional triangulation technique is not very reliable for locating the impact point even in isotropic plates when the sensors are placed close to the point of strike for two reasons: First, it is difficult to pinpoint the exact time of arrival of the signal and, second, the Lamb modes in a plate are dispersive. Dispersive signals attenuate differently at various frequencies and propagate with different speeds causing distortions in the received signals, and thus introduce error in the time of flight measurement. The triangulation technique assumes that wave speeds in all directions are the same, which is not true for anisotropic plates. Here an alternative approach based on an optimization scheme is proposed to locate the point of impact in isotropic and anisotropic plates. A formulation is presented for the general anisotropic case. Experiments are carried out with an aluminum plate by dropping balls on the plate and picking up acoustic signals at different locations. The impact points predicted by the conventional triangulation technique and the proposed modified method are compared for this isotropic plate. Then it is investigated how the prediction would change if the plate is assumed to have some anisotropy. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2775322]

I. INTRODUCTION

Nondestructive evaluation techniques using ultrasonic transducers can efficiently locate and characterize existing and emerging defects. Ultrasonic transducers are generally used in two modes commonly known as active and passive modes[1] for monitoring structural damage. Under active mode monitoring, acoustic actuators generate ultrasonic signals[2] and under passive mode monitoring the impact of foreign objects or crack initation/growth act as the acoustic source.[3,4] For both applications, ultrasonic sensors are placed in critical areas of the structure to efficiently receive ultrasonic signals and monitor its condition.[5–10]

In this paper attention is focused on the passive monitoring technique to locate the point of impact of a foreign object in a plate. For isotropic plates the point of impact can be located after detecting the acoustic emission signal (generated by the impact phenomenon) by at least three sensors and applying the triangulation technique. The triangulation technique works very well when the velocity of the signal (c) in the medium and the time of arrival (t_i) of the signal at all three sensor locations are known. The impact point is identified by drawing three circles of radii (r_i), whose centers coincide with the three sensor locations. The radius r_i is obtained by multiplying the time of arrival of the signal (t_i) with the signal velocity (c). The intersection point of these three circles is the point of impact. However, for dispersive guided waves (such as Lamb waves in a plate) the wave velocity is not constant but varies with the signal frequency. Signal frequency depends on the properties of the foreign object striking the structure. Therefore, when a plate is struck by an object the velocity of the generated Lamb mode may be unknown. Also, the exact time of arrival of the signal may not be known either because of the distortion of the signal due to dispersion or because the instrument noise level hiding the exact time of arrival of the signal. If the plate is anisotropic then also the triangulation technique does not work. This paper presents an alternative method to the triangulation technique; the proposed method works for both isotropic and anisotropic plates. This method is based on some optimization algorithm that minimizes an error function or objective function. It should work for isotropic plates even when the wave speed in the plate and the exact arrival times of the signal at the station locations are not known. For this technique to work for anisotropic plates the direction dependence of the wave velocity in the plate must be known. This technique is an improvement over the previously reported optimization-based techniques.[11,12] This technique does not need to exploit the directivity properties of piezoceramic sensors[13] and, therefore, the sensor can have any shape and size.

II. EXPERIMENT

Experiments were carried out by dropping two different balls on an aluminum plate. Thus the ball drop event simulates the impact phenomenon. Two balls used in the experiments are shown in Fig. 1—the bigger ball on the left is a ping pong ball and the smaller ball on the right is a soft

---

[a)]Electronic mail: tkundu@email.arizona.edu
[b)]Electronic mail: samik@email.arizona.edu
[c)]Electronic mail: kumar.jata@wpafb.af.mil
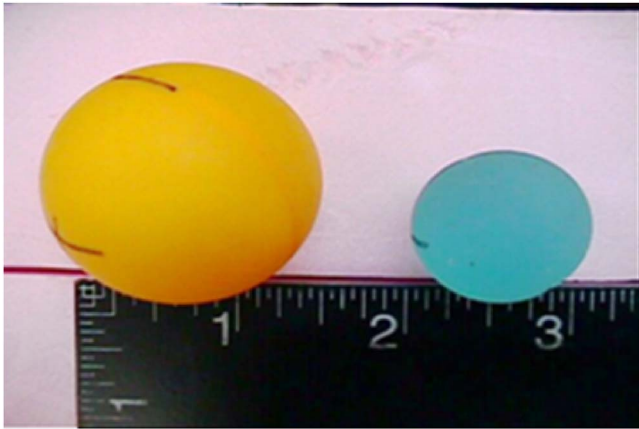
FIG. 1. (Color online) A ping pong ball (left) and a soft rubber ball (right) used in the ball-drop experiment. The scale shows dimensions in inches (1 in.=25.4 mm).

FIG. 3. (Color online) Photograph of the ball drop experimental setup. Vertical pipe above the inclined plate guides the ball drop path.

rubber ball. A relatively hard ping pong ball and soft rubber ball were used to generate elastic waves in the plate in two different frequency ranges.

To avoid the ball bouncing on the plate multiple times the plate was placed at a slightly inclined position. The inclined position of the plate helped the ball, after its first bounce, to move away from its vertical line of drop and the plate as shown in Fig. 2. For such plate orientation it was easy to catch the ball after its first bounce. Figure 3 shows how the plate was supported and the ball drop path was controlled. The aluminum plate was placed in a square tank with one side supported at an elevation. The ball was dropped through a vertical pipe placed above the plate as shown in Fig. 3. The ball drop height was varied between 7 in. (177.8 mm) and 24 in. (609.6 mm) by placing the ball on a cardboard inside the vertical pipe at different heights and removing the cardboard through the slit openings of the pipe. Slit openings at regular intervals of the pipe were cut (see Fig. 3) so that a card board could be placed horizontally inside the pipe through those slit openings to hold the ball. This arrangement accurately controlled the ball drop height and guaranteed that during repeated experiments the ball was striking the same location of the plate.

Four passive sensors were adhesively bonded to the plate. Closer views of these sensors mounted on the plate are shown in Fig. 4. The image on the left-hand side of Fig. 4 shows a piezoelectric acoustic sensor mounted away from the point of strike. The image on the right-hand side shows a much bigger PVDF sensor placed at the impact location.

Schematic diagram of the aluminum plate with mounted acoustic emission sensors is shown in Fig. 5. The plate dimension is 24 in. × 24 in. (61 cm × 61 cm) and its thickness is 4.11 mm. Sensor locations and the impact point are also shown in Fig. 5. S1, S2, and S3 are the locations of the three acoustic emission sensors that were placed away from the point of impact. White square denotes the point of impact, and the location of the fourth sensor is denoted as S0. Received signals at the four sensors (S1, S2, S3, and S0) are displayed on a four-channel oscilloscope screen.

III. THEORETICAL DEVELOPMENT

Let the time of detection of the acoustic signal at the ith sensor or ith station be denoted as t_{id}. If the time of impact is t_c then the travel time for the signal from the impact point to the station location is
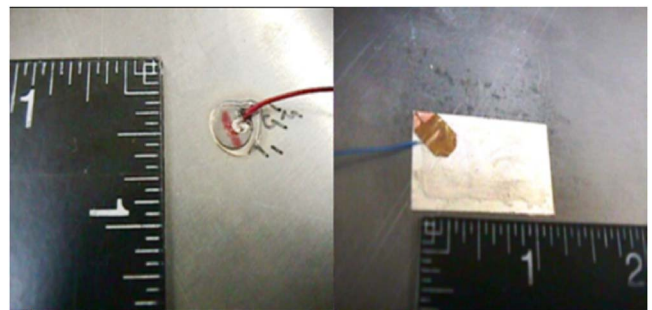
FIG. 4. (Color online) Acoustic emission sensors mounted on the plate—the left image is for a piezoelectric sensor mounted away from the point of impact and the right image shows a much bigger PVDF sensor mounted at the impact position. The scale shows dimensions in inches (1 in. =25.4 mm).

FIG. 2. (Color online) Schematic of the plate orientation for the ball-drop experiment.

Kundu *et al.*: Impact point prediction
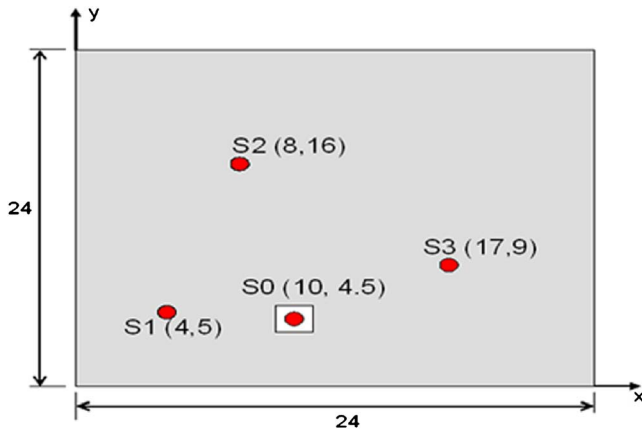
FIG. 5. (Color online) The plate specimen showing the point of impact, marked by the white square and four acoustic emission sensors denoted by four small circles. Three sensors S1, S2, and S3 are placed away from the impact point while the fourth sensor (S0) is placed at the impact position. Coordinates of the four sensors and the plate dimensions are given in inches (1 in.=25.4 mm).

$$t_i = t_{id} - t_c. \tag{1}$$

Note that in Eq. (1) both t_{id} and t_c are defined in the same time of reference.

If the coordinates of the three sensors S1, S2, and S3 are (x_1, y_1), (x_2, y_2), and (x_3, y_3), respectively, and the impact point coordinate is (x_0, y_0), then the distance of the three sensors from the impact point is given by

$$d_1 = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2},$$

$$d_2 = \sqrt{(x_2 - x_0)^2 + (y_2 - y_0)^2}, \tag{2}$$

$$d_3 = \sqrt{(x_3 - x_0)^2 + (y_3 - y_0)^2}.$$

The time of travel of the guided wave to the three sensor locations are denoted as t_1, t_2, and t_3, respectively, and the velocity $[c(\theta)]$ of the wave in the plate is a function of the wave propagation direction θ; therefore, one can write

$$d_1 = c(\theta_1)t_1 = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2},$$

$$d_2 = c(\theta_2)t_2 = \sqrt{(x_2 - x_0)^2 + (y_2 - y_0)^2}, \tag{3}$$

$$d_3 = c(\theta_3)t_3 = \sqrt{(x_3 - x_0)^2 + (y_3 - y_0)^2}.$$

From Eq. (3) one obtains

$$\frac{d_2}{d_1} = \frac{c(\theta_2)t_2}{c(\theta_1)t_1} = \frac{\sqrt{(x_2 - x_0)^2 + (y_2 - y_0)^2}}{\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}} = r_{21},$$

$$\frac{d_3}{d_1} = \frac{c(\theta_3)t_3}{c(\theta_1)t_1} = \frac{\sqrt{(x_3 - x_0)^2 + (y_3 - y_0)^2}}{\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}} = r_{31}.$$

$$\tag{4}$$

Therefore,

$$\left(\frac{d_2}{d_1}\right)^2 + \left(\frac{d_3}{d_1}\right)^2 = \left(\frac{c(\theta_2)t_2}{c(\theta_1)t_1}\right)^2 + \left(\frac{c(\theta_3)t_3}{c(\theta_1)t_1}\right)^2 = \frac{(x_2 - x_0)^2 + (y_2 - y_0)^2}{(x_1 - x_0)^2 + (y_1 - y_0)^2} + \frac{(x_3 - x_0)^2 + (y_3 - y_0)^2}{(x_1 - x_0)^2 + (y_1 - y_0)^2}$$

$$\Rightarrow r_{21}^2 + r_{31}^2 = \frac{(x_2 - x_0)^2 + (x_3 - x_0)^2 + (y_2 - y_0)^2 + (y_3 - y_0)^2}{(x_1 - x_0)^2 + (y_1 - y_0)^2}$$

$$\Rightarrow r_{21}^2 + r_{31}^2 = \frac{2(x_0^2 + y_0^2 - x_0 x_2 - x_0 x_3 - y_0 y_2 - y_0 y_3) + x_2^2 + x_3^2 + y_2^2 + y_3^2}{x_0^2 + y_0^2 - 2(x_0 x_1 + y_0 y_1) + x_1^2 + y_1^2}. \tag{5}$$

From Eq. (5) it is possible to define an error function $E(x_0, y_0)$, also known as the objective function in the optimization literature, in the following manner:

$$E(x_0, y_0) = [2(x_0^2 + y_0^2 - x_0 x_2 - x_0 x_3 - y_0 y_2 - y_0 y_3)$$
$$+ x_2^2 + x_3^2 + y_2^2 + y_3^2 - (r_{21}^2 + r_{31}^2)$$
$$\times (x_0^2 + y_0^2 - 2(x_0 x_1 + y_0 y_1) + x_1^2 + y_1^2)]^2. \tag{6}$$

It should be noted here that during real life structural health monitoring applications the sensor locations (x_1, y_1), (x_2, y_2), and (x_3, y_3) are known, but not the impact point (x_0, y_0). If the coordinate values (x_0, y_0) are assumed for the impact point and that assumption is wrong, then a positive value of the error function $E(x_0, y_0)$ of Eq. (6) should be obtained. However, if the assumption is correct then the error function should give a zero value. Therefore, one can obtain (x_0, y_0) by minimizing the error function value.

IV. EXPERIMENTAL RESULTS

In Fig. 5 placing the origin of the xy-coordinate system at the bottom-left corner of the plate and measuring the x coordinate from the left to the right and the y coordinate from the bottom to the top one gets the coordinate values in inches (1 in.=25.4 mm) as shown in Fig. 5 and given as follows:

$$x_1 = 4 \text{ in.,} \quad y_1 = 5 \text{ in.,}$$

$$x_2 = 8 \text{ in.,} \quad y_2 = 16 \text{ in.,}$$

$$x_3 = 17 \text{ in.,} \quad y_3 = 9 \text{ in.,}$$

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Kundu *et al.*: Impact point prediction   2059
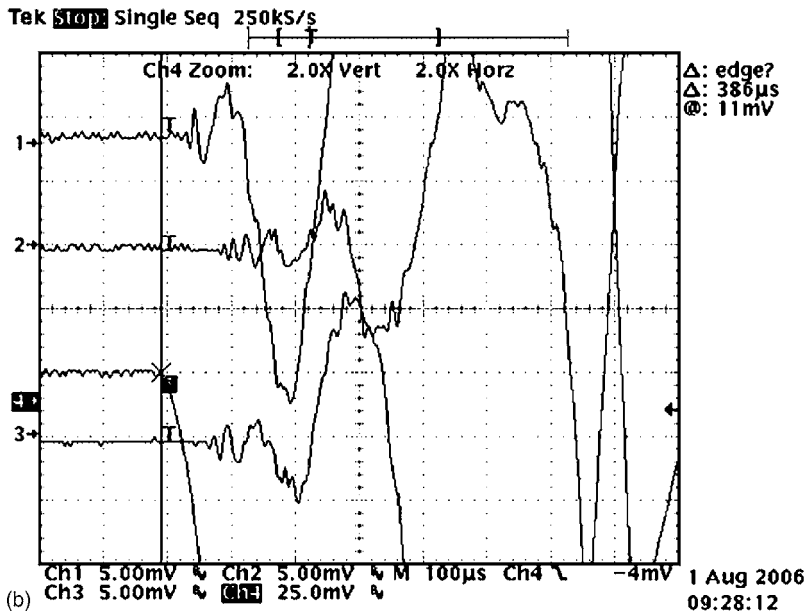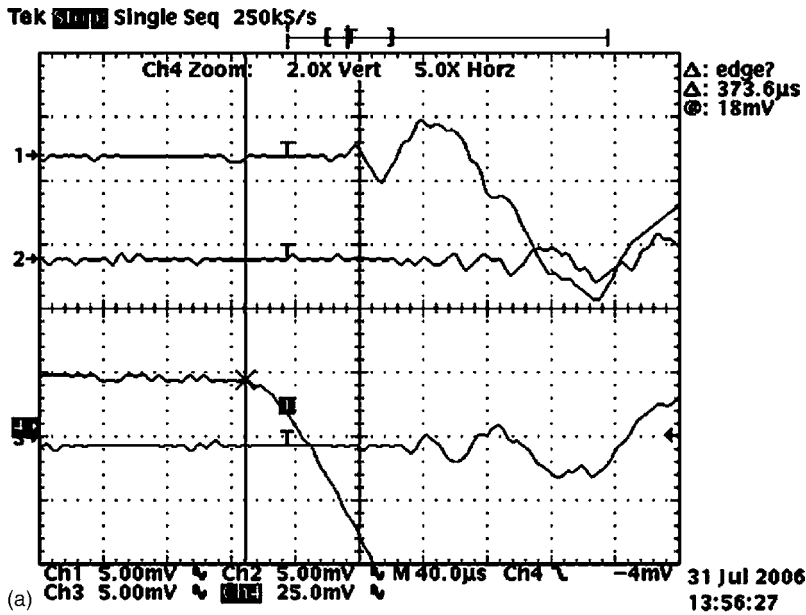
FIG. 6. (a) For the rubber ball drop experiment the time histories recorded at three receiving stations (S1, S2, and S3, corresponding curves are marked as 1, 2, and 3, respectively) and by the sensor placed at the impact point (S0, marked as curve 4). (b) For the ping pong ball drop experiment the time histories recorded at three receiving stations (S1, S2, and S3, corresponding curves are marked as 1, 2, and 3, respectively) and by the sensor placed at the impact point (S0, marked as curve 4).

$$x_0 = 10 \text{ in.}, \quad y_0 = 4.5 \text{ in.}$$

Then

$$d_1 = \sqrt{(4-10)^2 + (5-4.5)^2} = \sqrt{37} = 6.08 \text{ in.},$$

$$d_2 = \sqrt{(8-10)^2 + (16-4.5)^2} = 11.67 \text{ in.},$$

$$d_3 = \sqrt{(17-10)^2 + (9-4.5)^2} = 8.32 \text{ in.}$$

As mentioned before, a soft rubber ball and a ping pong ball were dropped from different heights, varying between 7 in. (177.8 mm) and 24 in. (609.6 mm) at the same location (x_0, y_0) of the plate. Time histories were recorded at four sensor positions for several impact experiments with the two balls. Typical time histories and their Fourier transforms are shown in Figs. 6(a), 6(b), 7, and 8. It should be noted here that only the initial parts of the time history plots are of interest for obtaining the arrival times of the guided waves at

the sensor locations. Spectral plots (Figs. 7 and 8) show the frequency contents of the arriving signals. These plots help one to identify the guided wave modes that are generated in the plate.

From the time history of the fourth sensor (placed at the impact point and denoted as S0 in Fig. 5) the exact time of impact can be obtained. The impact time has been denoted by a solid vertical straight line in Figs. 6(a) and 6(b). In Fig. 6(a), 1 cm in the horizontal direction (the length between two neighboring vertical dashed lines) corresponds to 40 μs, while in Fig. 6(b) the same length corresponds to 100 μs. It is not easy to measure the exact time of arrival of the signals from the time history plots because there is some ambiguity in these plots about the exact arrival time or the starting point of the signals. Note that there is always some noise in the time history plots before the arrival of the ultrasonic energy. The exact arrival time is hidden in this noise. Therefore, depending on which point we consider as the starting point

2060   J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Kundu *et al.*: Impact point prediction
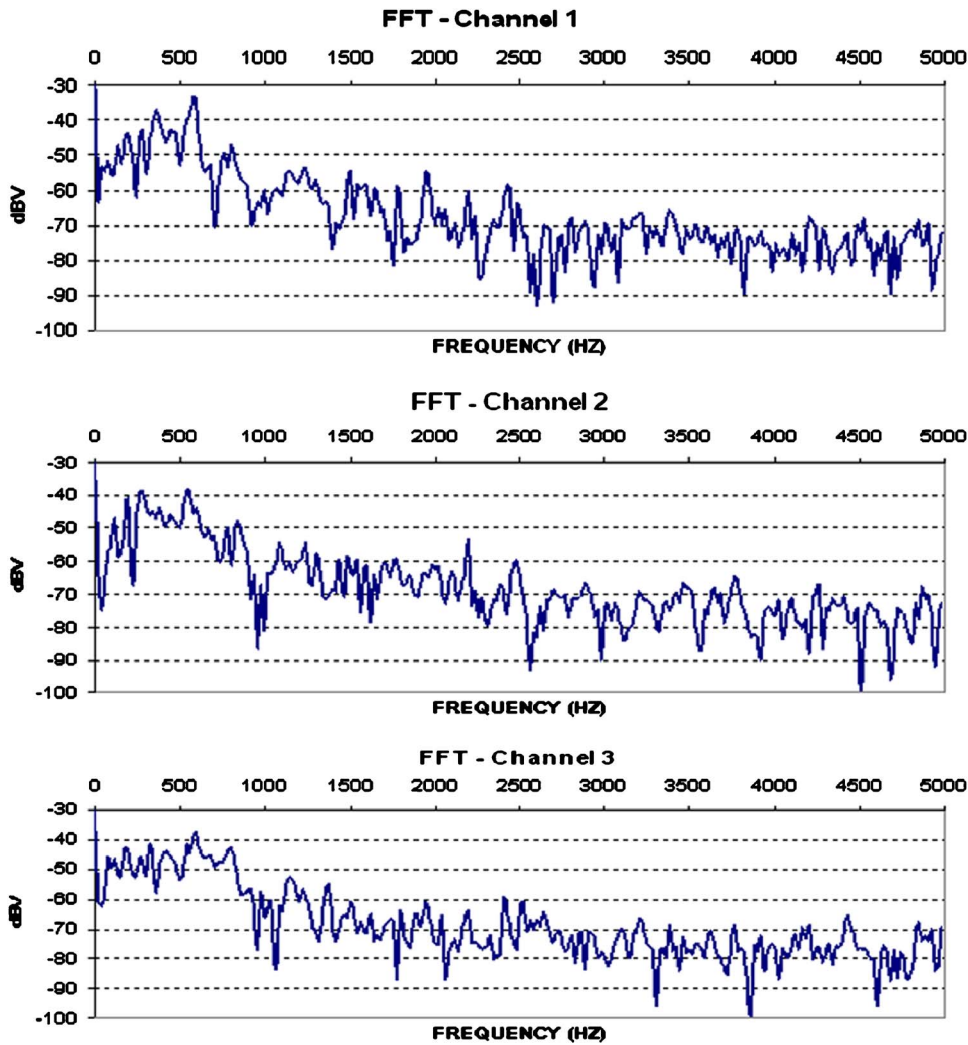
FIG. 7. (Color online) Fourier transforms of three time histories for the rubber ball drop experiment.

of the received signal, different values for the time of arrival are obtained. For example, if we consider the first noticeable peak above the noise level as arrival times for the three signals, then arrival times are 1.80, 3.37, and 2.89×40 μs. Note that the ratio of the three arrival times is 1:1.87:1.61, while the distance ratio $(d_1 : d_2 : d_3)$ from the point of impact is 1:1.92:1.37. A number of such arrival time ratios from multiple experimental time history plots were computed and similar type of discrepancies between the arrival time ratio $(t_1 : t_2 : t_3)$ and the actual distance ratio $(d_1 : d_2 : d_3)$ were observed. The point of impact predicted by the triangulation technique for this arrival time ratio is shown in Fig. 9 by the triangular marker. Note that this point is more than 1 in. (2.54 cm) apart from the actual point of impact. If the impact point is predicted from the time history signals generated by the ping pong ball [Fig. 6(b)] then the error increases even more. In Fig. 9 one can clearly see that the difference between the actual impact point and the predicted impact point is much greater for the ping pong ball result (square marker).

From the frequency domain spectral plots of Figs. 7 and 8 one can see that for the rubber ball drop experiment the frequency content above −60 dB cut off line (−60 dB is about the noise level) is mostly between 0 and 1000 Hz, while the peak frequency is occurring between 200 and 800 Hz. For the ping pong ball drop experiment the signal above −60 dB cut off line is mostly between 0 and 2700 Hz, while the peak frequency is occurring between 800 and 1600 Hz. To identify the Lamb modes corresponding to these frequencies the dispersion curves for the aluminum plate are generated and shown in Fig. 10. For our experiment the plate thickness is 4.11 mm and the dominant frequency is between 0.2 and 1.6 kHz. Therefore the thickness times the frequency varies between 0.8 and 6.4 kHz mm or between 0.0008 to 0.0064 MHz mm. For such small values of the frequency-thickness product the S_0 mode has a velocity close to 5.5 mm/μs and A_0 mode has a very small velocity, close to zero.

A. Difficulties associated with the time of arrival measurement

The times of arrival as measured by the first detectable positive peaks in Fig. 6(a) are 1.80, 3.37, and 2.89×40 μs. For S_0 speed of 5.5 mm/μs these three arrival times correspond to 39.60, 74.14, and 63.58 cm, respectively. Clearly these values are much greater than the actual distances, which are

$$d_1 = 6.08 \text{ in.} = 15.44 \text{ cm},$$

$$d_2 = 11.67 \text{ in.} = 29.64 \text{ cm},$$
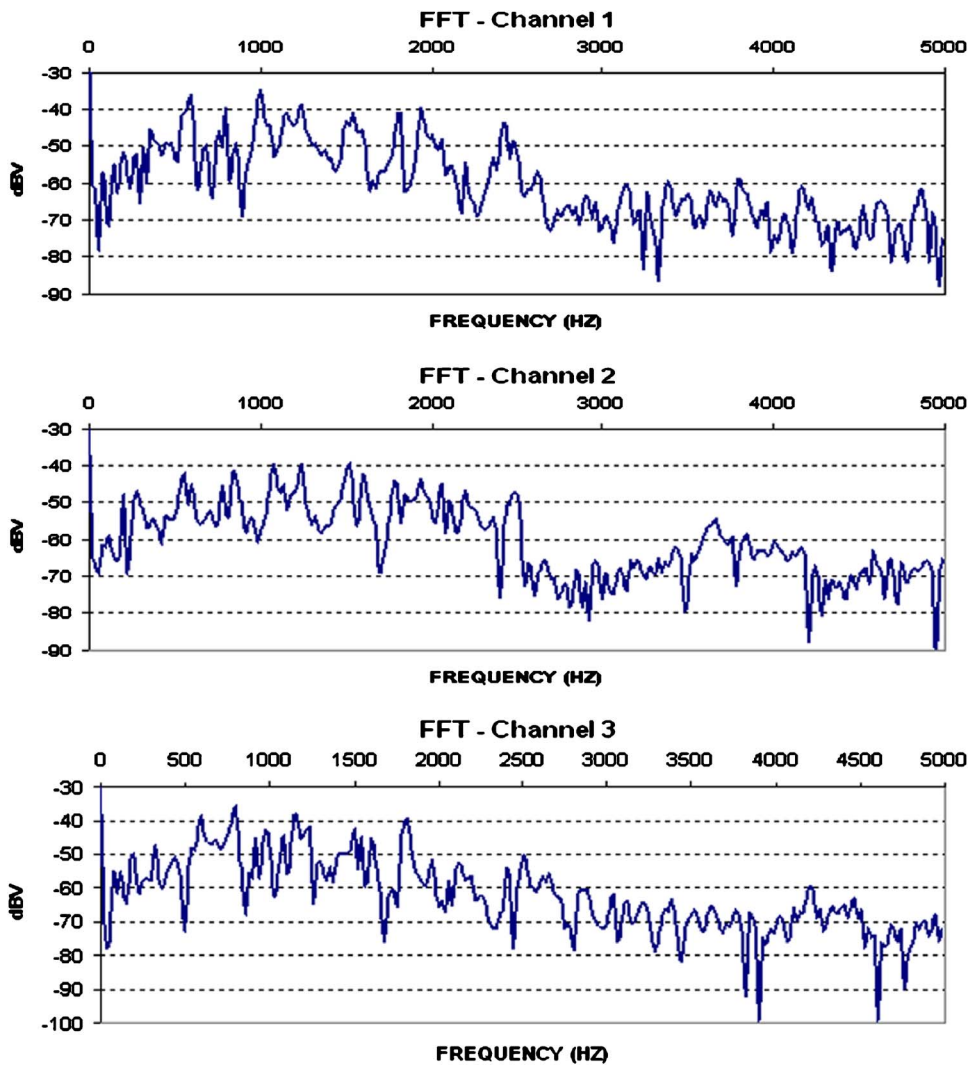
FIG. 8. (Color online) Fourier transforms of three time histories for the ping pong ball drop experiment.

$d_3 = 8.32$ in. $= 21.13$ cm.

Therefore, we see that the arrival time of the S_0 mode is much before the recorded time for the clearly detectable peak and is difficult to pinpoint because it is hidden in the noise level. Moreover, in reality the point of impact may not coincide with any sensor location and thus the exact time of impact is generally an unknown quantity. Therefore, even if the time of arrival of a signal at a source location [t_{id} in Eq.

(1)] can be accurately recorded we cannot determine the time of travel of the signal [t_i in Eq. (1)] without knowing the correct time of impact [t_c in Eq. (1)]. Therefore, it is necessary to define an alternate error function that does not require knowledge of the absolute arrival times—t_1, t_2, and t_3. The modified error function should take as input the difference in the recorded arrival times $t_{21}(=t_{2d}-t_{1d})$, $t_{31}(=t_{3d}-t_{1d})$, etc. It should be noted here that the difference in the recorded ar-
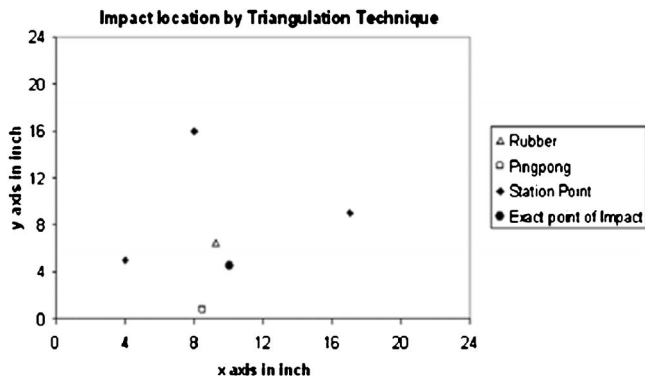
FIG. 9. Optimized location of point of impact using conventional triangulation technique.
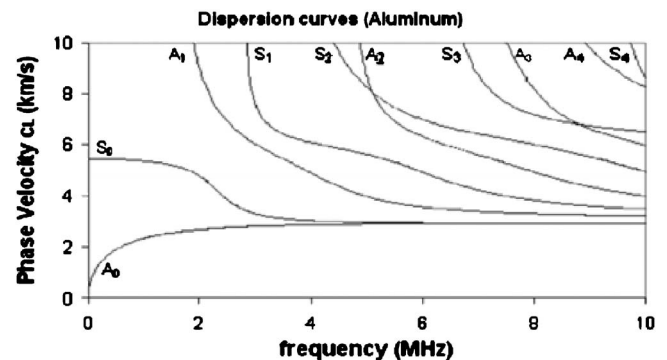
FIG. 10. Dispersion curves for 1-mm-thick aluminum plate. For aluminum plates of different thickness, one can use this plot by simply changing the horizontal scale to MHz mm (frequency × plate thickness).

rival times is the same as the difference in the actual times of travel as illustrated in the following:

$$t_{ij} = t_{id} - t_{jd} = (t_i + t_c) - (t_j + t_c) = t_i - t_j. \tag{7}$$

The new error functions are derived in Sec. IV B.

B. Alternate derivation of error function or objective function

If the wave speed is $c(\theta)$ then

$$t_{21} = \frac{d_2}{c(\theta_2)} - \frac{d_1}{c(\theta_1)} = \frac{c(\theta_1)d_2 - c(\theta_2)d_1}{c(\theta_2)c(\theta_1)} = \frac{1}{c(\theta_2)c(\theta_1)}\{c(\theta_1)\sqrt{(x_2 - x_0)^2 + (y_2 - y_0)^2} - c(\theta_2)\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}\}, \tag{8}$$

$$t_{31} = \frac{c(\theta_1)d_3 - c(\theta_3)d_1}{c(\theta_3)c(\theta_1)} = \frac{1}{c(\theta_3)c(\theta_1)}\{c(\theta_1)\sqrt{(x_3 - x_0)^2 + (y_3 - y_0)^2} - c(\theta_3)\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}\}$$

$$\Rightarrow \frac{t_{21}}{t_{31}} = \frac{c(\theta_3)}{c(\theta_2)} \frac{\{c(\theta_1)\sqrt{(x_2 - x_0)^2 + (y_2 - y_0)^2} - c(\theta_2)\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}\}}{\{c(\theta_1)\sqrt{(x_3 - x_0)^2 + (y_3 - y_0)^2} - c(\theta_3)\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}\}}.$$

Error function $E(x_0, y_0)$ can be defined as

$$E(x_0, y_0) = \left(\frac{c(\theta_3)}{c(\theta_2)} \frac{\{c(\theta_1)\sqrt{(x_2 - x_0)^2 + (y_2 - y_0)^2} - c(\theta_2)\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}\}}{\{c(\theta_1)\sqrt{(x_3 - x_0)^2 + (y_3 - y_0)^2} - c(\theta_3)\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}\}} - \frac{t_{21}}{t_{31}} \right)^2. \tag{9}$$

Ideally, for the correct values of (x_0, y_0) the error function should give a zero value while for wrong values of (x_0, y_0) it should give a positive value. Therefore, we need to minimize the value of this error function, which is also known as the objective function in the optimization literature. Note that in the above-presented definition of the error function the time t_1 has been used twice in computing t_{21} and t_{31} while t_2 and t_3 have been used only once. To give equal importance or bias to the three measurements of arrival times at the three sensor locations the error function can be defined in a different manner as shown in the following. With this definition of the error function, the impact point location prediction should not be strongly influenced by the experimental error in any one time of arrival measurement,

$$E(x_0, y_0) = \left(\frac{c(\theta_3)}{c(\theta_1)} \left\{ \frac{c(\theta_2)\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} - c(\theta_1)\sqrt{(x_2 - x_0)^2 + (y_2 - y_0)^2}}{c(\theta_3)\sqrt{(x_2 - x_0)^2 + (y_2 - y_0)^2} - c(\theta_2)\sqrt{(x_3 - x_0)^2 + (y_3 - y_0)^2}} \right\} - \frac{t_{12}}{t_{23}} \right)^2$$

$$+ \left(\frac{c(\theta_1)}{c(\theta_2)} \left\{ \frac{c(\theta_3)\sqrt{(x_2 - x_0)^2 + (y_2 - y_0)^2} - c(\theta_2)\sqrt{(x_3 - x_0)^2 + (y_3 - y_0)^2}}{c(\theta_1)\sqrt{(x_3 - x_0)^2 + (y_3 - y_0)^2} - c(\theta_3)\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}} \right\} - \frac{t_{23}}{t_{31}} \right)^2$$

$$+ \left(\frac{c(\theta_2)}{c(\theta_3)} \left\{ \frac{c(\theta_1)\sqrt{(x_3 - x_0)^2 + (y_3 - y_0)^2} - c(\theta_3)\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}}{c(\theta_2)\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} - c(\theta_1)\sqrt{(x_2 - x_0)^2 + (y_2 - y_0)^2}} \right\} - \frac{t_{31}}{t_{12}} \right)^2. \tag{10}$$

By some optimization scheme such as the simplex algorithm the impact point (x_0, y_0) can be obtained by minimizing the above-noted error function. It should be mentioned here that in this definition of the error function the differences in the arrival times are used as input, hence it is not necessary to know the exact time of impact.

The predictions can be improved further by placing more receiving sensors on the structure. For n number of sensors there should be n terms in the error function as defined in the following:

$$E(x_0, y_0) = (d_{123} - t_{123})^2 + (d_{234} - t_{234})^2 + (d_{345} - t_{345})^2$$

$$+ \cdots + (d_{(n-2)(n-1)n} - t_{(n-2)(n-1)n})^2$$

$$+ (d_{(n-1)n1} - t_{(n-1)n1})^2 + (d_{n12} - t_{n12})^2, \tag{11}$$

where

$$t_{ijk} = \frac{t_{ij}}{t_{jk}} = \frac{t_i - t_j}{t_j - t_k} \tag{12}$$

and

$$d_{ijk} = \frac{c(\theta_k)}{c(\theta_i)} \left\{ \frac{c(\theta_j)\sqrt{(x_i-x_0)^2+(y_i-y_0)^2} - c(\theta_i)\sqrt{(x_j-x_0)^2+(y_j-y_0)^2}}{c(\theta_k)\sqrt{(x_j-x_0)^2+(y_j-y_0)^2} - c(\theta_j)\sqrt{(x_k-x_0)^2+(y_k-y_0)^2}} \right\}. \tag{13}$$

Note that the angle of the wave propagation direction from the source (x_0, y_0) to the station (x_i, y_i) is measured from the horizontal axis and can be obtained from

$$\theta_i = \tan^{-1}\left(\frac{y_i - y_0}{x_i - x_0} \right). \tag{14}$$

Equation (14) is valid for all possible combinations of (x_0, y_0) and (x_i, y_i) for which the computed θ_i values should vary between $-\pi/2$ and $+\pi/2$. Since the wave velocity in θ_i and $(\theta_i + \pi)$ directions should be the same it is not necessary to consider any angle beyond the boundary $(-\pi/2$ and $+\pi/2)$ for computing the wave velocity in all possible directions between $-\pi/2$ and $+3\pi/2$.

V. SIGNAL PROCESSING FOR PREDICTING THE POINT OF IMPACT

For predicting the impact location with the above-described experimental setup the error function or the objective function given in Eq. (10) has to be minimized. For an alternate experimental setup consisting of $n(>3)$ number of sensors the objective function in Eq. (11) has to be minimized. As mentioned before, for both these equations it is not necessary to know the exact time of travel of the ultrasonic signal from the source to the station.

Error function given in Eq. (10) is plotted in Fig. 11 for an isotropic plate for which $c(\theta_i)$ from numerator and denominator cancels. In Fig. 11 both x_0 and y_0 values vary from 0 to 24 in. Several sharp peaks are observed in this plot because at those points the error function becomes singular (goes to infinity). Note that when x_0 and y_0 values are taken such that the point (x_0, y_0) is equidistant from any two sensor positions then the error function becomes singular. For example, when (x_0, y_0) is equidistant from (x_1, y_1) and (x_2, y_2), then the denominator of the third term of $E(x_0, y_0)$ defined in
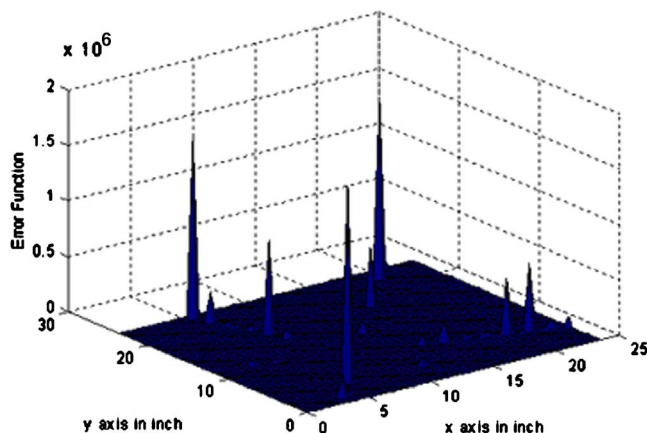
Eq. (10) becomes zero for an isotropic plate making the error function value infinity. Therefore, for an isotropic plate the error function value becomes infinity on the perpendicular bisector of the line joining (x_1, y_1) and (x_2, y_2) points. Similarly, it will be infinity on the perpendicular bisectors of the lines joining $(x_1, y_1) - (x_3, y_3)$ points and $(x_2, y_2) - (x_3, y_3)$ points. Therefore, the error function of Eq. (10) should be singular on three straight lines. High values of the error function along these three straight lines can be seen in Figs. 11 and 12. In Fig. 12, logarithm of the error function is plotted. Because of the large values of the error function near the singular points, the function looks almost flat at other points. However, when logarithm of the error function is plotted in Fig. 12, a better variation of the error function can be seen.

It is well known that the SIMPLEX algorithm[14] and many other optimization schemes fail if an objective function has a large number of local minima. For this problem the SIMPLEX algorithm was tried first but because of a large number of local minima it did not converge to the right values. In different trials the method converged to different coordinates as shown in Fig. 13. Clearly, these predictions are worse than the results obtained by the conventional triangulation technique (Fig. 9).

To improve the prediction a modified technique was sought. In this technique the global minimum of the error function was captured by computing its value for a large number of x and y values after constructing a fine mesh grid on the plate dimension. With the help of MATLAB the minimum value of all these computed values at the mesh grid points was captured and corresponding coordinates for this minimum error function value were noted. First, x coordinate is varied from 0 to 24 and the minimum value of the error function for each x value is plotted as shown in Fig. 14(a). Note that the smallest value is noted for $x=8.56$. Then y coordinate is changed from 0 to 24 and the minimum value of the error function for each y value is plotted as shown in Fig. 14(b). The smallest value is obtained for $y=3.81$. There-

FIG. 11. (Color online) Plot of the error function [Eq. (10)] for the time of arrival values obtained from the top plot of Fig. 6.
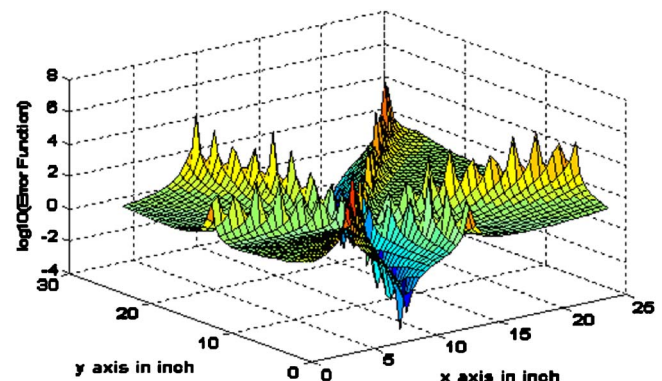
FIG. 12. (Color online) Plot of the logarithm of the error function of Eq. (10) for the time of arrival values obtained from the top plot of Fig. 6.
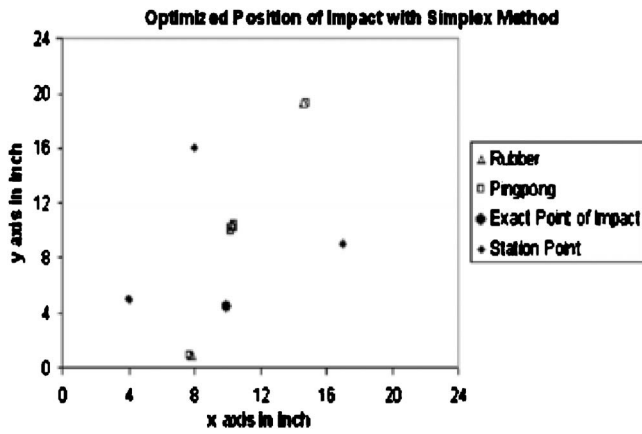
FIG. 13. Predicted points of impact using SIMPLEX algorithm for different initial estimates. Note that most predictions (open triangles and open squares) are far away from the exact point of impact (closed circle).

fore, the predicted impact point coordinate is (8.56, 3.81). In a similar manner, impact point coordinates from all other experimental readings were computed. These coordinates are shown in Fig. 15. The predictions of the impact point in Fig. 15 are better than those in Figs. 9 and 13. The optimization scheme based on the genetic algorithm also produced results similar to Fig. 15.
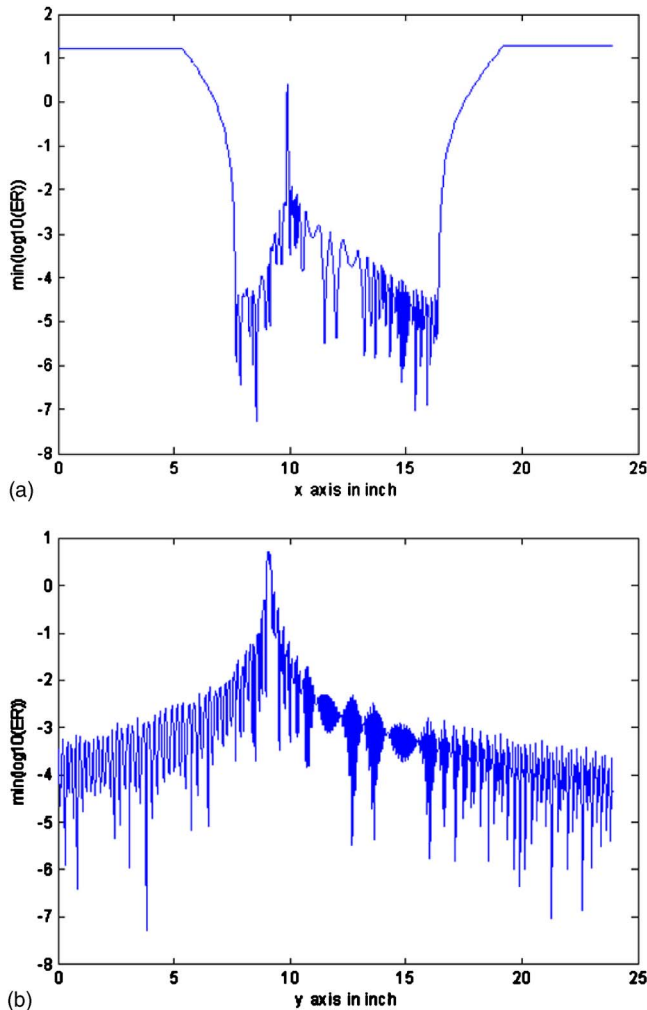
FIG. 14. (Color online) Plot of minimum of logarithm of error function against (a) the x-axis coordinates and (b) the y-axis coordinates.
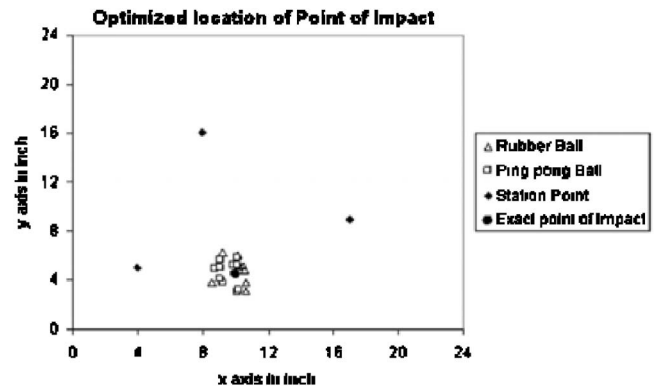
FIG. 15. Predicted points of impact (open triangles and open squares) from the fine mesh grid analysis.

VI. PREDICTING THE POINT OF IMPACT ASSUMING ANISOTROPIC PLATE PROPERTIES

Often, small degrees of anisotropy in cold-rolled metal plates are observed. If we assume 1% increase in the wave velocity for every 10° change in the wave propagation direction from the horizontal axis then the wave velocity as a function of the wave propagation direction becomes

$$c(\theta) = c_0(1 + 0.001|\theta|). \tag{15}$$

In Eq. (15) θ is in degrees and c_0 is the wave velocity in the horizontal direction. Impact point predictions for this anisotropic plate with the same set of experimental results are shown in Fig. 16. From Figs. 15 and 16 one can see that for small degree of anisotropy in the plate material the impact point predictions do not change significantly. Therefore, for such small anisotropy it is acceptable to assume the plate to be isotropic for impact point prediction.

VII. CONCLUSION

A new technique to locate the point of impact of a foreign object on a plate is proposed. The proposed technique is based on minimizing an objective function or error function that does not require the knowledge of the exact time of travel of the signal from the source to the station. This objective function uses the difference in time of arrival of the ultrasonic signals instead of the absolute time of arrival of the signal. Experimental results show the superiority of this technique over the conventional triangulation technique that
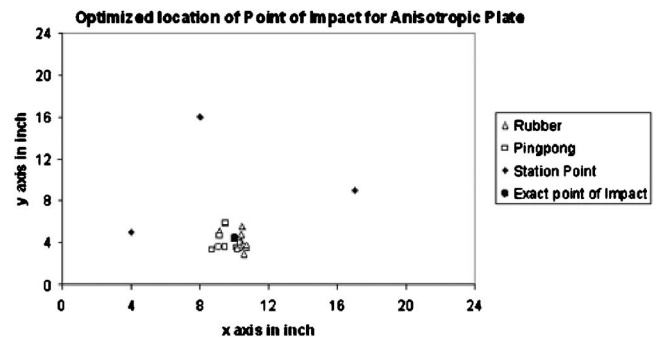
FIG. 16. Predicted points of impact (open triangles and open squares) from the fine mesh grid analysis for the anisotropic plate, Eq. (15) gives the degree of anisotropy.

works for isotropic media only. The proposed technique works equally well for isotropic and anisotropic plates. By locating the global minimum of the objective function the convergence problem of the SIMPLEX algorithm to local minima has been avoided. This research is important for structural health monitoring of isotropic and anisotropic plate structures.

ACKNOWLEDGMENTS

[1] K. V. Jata, T. Kundu, and T. Parthasarathy, "An introduction to failure mechanisms and ultrasonic inspection," in *Advanced Ultrasonic Methods for Material and Structure Inspection*, edited by T. Kundu (ISTE, London, 2007), Chap. 1, pp. 1–42.

[2] V. Giurgiutiu, "Lamb wave generation with piezoelectric wafer active sensors for structural health monitoring," Proc. SPIE **5056**, 111–122 (2003).

[3] A. K. Mal, F. Shih, and S. Banerjee, "Acoustic emission waveforms in composite laminates under low velocity impact," Proc. SPIE **5047**, 1–12 (2003).

[4] A. K. Mal, F. Ricci, S. Gibson, and S. Banerjee, "Damage detection in structures from vibration and wave propagation data," Proc. SPIE **5047**, 202–210 (2003).

[5] B. Köhler, F. Schubert, and B. Frankenstein, "Numerical and experimental investigation of Lamb wave excitation, propagation and detection for SHM," *Proceedings of the Second European Workshop on Structural Health Monitoring*, Munich, Germany, 2004, pp. 993–1000.

[6] J. Park and F. K. Chang, "Built-in detection of impact damage in multi-layered thick composite structures," *Proceedings of the Fourth International Workshop on Structural Health Monitoring*, Stanford Univ., CA, USA, 2003, pp. 1391–1398.

[7] A. K. Mal, F. Ricci, S. Banerjee, and F. Shih, "A conceptual structural health monitoring system based on wave propagation and modal data," Struct. Health Monit. **4**, 283–293 (2005).

[8] G. Manson, K. Worden, and D. J. Allman, "Experimental validation of structural health monitoring methodology. II. Novelty detection on an aircraft wing," J. Sound Vib. **259**, 345–363 (2003).

[9] S. C. Wang and F.-K. Chang, "Diagnosis of impact damage in composite structures with built-in piezoelectrics network," Proc. SPIE **3990**, 13–19 (2000).

[10] S. S. Kessler, S. M. Spearing, and C. Soutis, "Damage detection in composite materials using Lamb wave methods," Smart Mater. Struct. **12**, 795–803 (2002).

[11] T. Kundu, S. Das, and K. V. Jata, "An improved technique for locating the point of impact from the acoustic emission data," Proc. SPIE **6532**, 1-12 (2007).

[12] P. T. Coverley and W. J. Staszewski, "Impact damage location in composite structures using optimized sensor triangulation procedure," Smart Mater. Struct. **12**, 795–803 (2003).

[13] F. L. Di Scalea, H. Matt, I. Bartoli, A. Srivastava, G. Park, and C. Farrar, "The fundamental response of piezoelectric guided-wave sensors and applications to damage and impact location," *Proceedings of the Third European Conference on Structural Health Monitoring*, Granada, Spain, 2006, pp. 1271–1278.

[14] J. A. Nelder and R. Mead, "A simplex method for function minimization," Comput. J. **7**, 308–315 (1965).

Approximate reconstruction of sound fields close to the source surface using spherical nearfield acoustical holography

Jens Prager[a)]

Institute of Fluid Mechanics and Engineering Acoustics, Technical University of Berlin, Berlin, Germany

This paper presents an investigation of the reconstruction of sound field parameters close to the surface of arbitrarily shaped sound sources. The field is reconstructed using nearfield acoustical holography (NAH) in spherical coordinates. Of particular interest are source shapes where the Rayleigh hypothesis is violated. To overcome the limitation of the minimal sphere given by the validity restriction of the Rayleigh hypothesis an algorithm is proposed for extracting local information from the nonconvergent NAH solution. For the assessment of the results an appropriate virtual test rig is developed employing the Kirchhoff–Helmholtz integral theorem. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2770539]

I. INTRODUCTION

The analysis and characterization of sources of sound and vibration remains a challenging area of research. In this pursuit, simplifications are required to make proposed methods applicable in engineering practice. As shown in recent textbooks, the methods of weighted residuals offer appropriate tools for source characterization. The basic idea is the expansion of the sound field in terms of outgoing wave components radiated from a set of discrete theoretical sources always located inside the vibrating body. Those sets of theoretical sources, e.g., a number of distributed monopole and/or dipole sources, establish the field at a predefined, continuous interface, simulating that of the actual source sufficiently accurately.[1] Alternatively, this is achieved by a single multipole. There the pseudosources of different order are concentrated in one point. This model can be seen as an analogy to a wave field expansion in spherical coordinates. Although the expansion is infinite in theory it is truncated in any practical realization and, hence, errors are introduced. The expansion coefficients can be considered as the complex strengths of the pseudo sources (poles). By knowing the expansion coefficients the field can be reconstructed on any point where the expansion is valid.

One possibility to determine the expansion coefficients is the nearfield acoustical holography (NAH). A goal of NAH in the given context can be to reconstruct the velocity of the vibrating source surface by simply measuring the sound pressure on a hypothetical surface at an accessible distance from the source. Because of the advantage of the closed, compact analysis surface, the NAH in spherical coordinates is desirable for this analysis. The periodicity of the analysis surface removes the requirement of spatial windowing and introduces no low frequency limit. Further, it provides a way to measure the source as one entity rather than at different time steps. This might be pivotal for transient processes. For the spherical NAH, however, it is required that

the sound source is completely enclosed by the analysis surface. This condition can be easily fulfilled if the shape of the source is nearly spherical. The velocity can be calculated directly on or close to the source surface. In engineering practice, the sound sources usually have complicated shapes. For arbitrarily shaped sources the limitation of the minimal sphere becomes important.

Consider the radiating body as a bounded domain $D \subset \mathbb{R}^3$, bounded by the surface Γ. Denote the exterior domain $D' = \mathbb{R}^3 \backslash D$. For all spheres B with the origin inside D completely enclosing D and accordingly not intercepting it at any point, the NAH approach is always valid and the field can be reconstructed for any point on the sphere. It should be mentioned that the field is represented entirely in terms of outgoing wave functions.

One century ago, Rayleigh[2] conjectured that the representation can also be applied unrestricted to the interior of B. With that, the wave field could also be predicted at the surface of any nonspherical source. At present, there is no doubt that the so-called Rayleigh hypothesis (or Rayleigh conjecture) does not generally apply to arbitrarily shaped source configurations. Millar[3] and van den Berg[4] showed that the validity of the Rayleigh hypothesis depends on whether or not there are singularities in the wave field. Consider a minimal sphere B_1 with smaller radius than B but the same origin so that all singularities of the sound field are inside B_1. Everywhere outside B_1 the Rayleigh hypothesis holds. Only if B_1 is completely inside D can the solution of the expansion be analytically continued to the entire source surface. Consequently, the hypothesis never holds if the surface has edges.[4,5] And this, however, is in general the case for practically relevant sources. This means that NAH is not applicable in engineering practice unrestrictedly.

To overcome this problem, Wu[6] suggested the Helmholtz equation-least-squares (HELS) method. The applicability of this method was demonstrated for simply shaped vibrating bodies with smooth surfaces. Alternatively, the Helmholtz integral theory can be used to form a NAH description.[7] This boundary elements method (BEM)-based NAH has been used successfully to reconstruct acoustic ra-

---

a)Electronic mail: jens.prager@tu-berlin.de

diation from structures.[8] Wu[9] employed a combination of both methods in the CHELS method. The outer domain between the measurement surface and the minimal sphere is solved with sound field reconstruction, whereas the inner domain is calculated using a BEM approach. For complex structures the number of nodes necessary to describe the acoustic field can be large, which makes the reconstruction process very time consuming.

The problem of sound field reconstruction is closely related to the handling of scattering by obstacles. In the publications on the modified Rayleigh conjecture, Ramm[10] has shown the sound field to converge overall on the surface of obstacles not fulfilling the Rayleigh hypothesis. With this in mind the question arises if there is an approximation of the sound field values directly on or close to the surface of arbitrary shaped vibrating bodies.

Hence, the focus of the work is on investigating the influence of the singularities in the sound field beside the commonly known fact that the field expansion is not uniformly convergent. With this knowledge a method will be proposed for an approximate prediction of the sound field not fulfilling the Rayleigh hypothesis. In this method the expansion is transformed into one that converges rapidly.

For the investigations a manageable virtual test rig is required. Such a test rig consists of a known source. It enables the determination of sound field parameters at any virtual surface surrounding the source. On that surface the sound field can be sampled. Since it is known that the NAH reconstruction of the sound field is sensitive to the presence of noise it will be added artificially to the sampled data. In Sec. II a virtual test rig is developed employing the Kirchhoff–Helmholtz integral (KHI).

II. MODELING THE TEST SOURCE

The virtual test rig developed in this section will be used in the same configuration for all investigations presented in this paper. The test rig is used for simulating the sound pressure on the measurement grid and for the assessment of the NAH results. Using the KHI, the sound pressure at any point in the exterior field of a source can be calculated by integrating the monopole and dipole parts of any surface element of the source. For harmonic motion, the KHI is given by[11]

$$p(\vec{x}_0) = \frac{i\omega\rho_0}{4\pi}\int_S \frac{e^{-ikr}}{r}v_n dS + \frac{1}{4\pi}\int_S p\frac{\partial}{\partial n}\left(\frac{e^{-ikr}}{r}\right)dS. \quad (1)$$

Herein n is the normal vector to the surface S of the source and $r=|\vec{x}_S-\vec{x}_0|$ the distance between the surface position and the observation point $\vec{x}_0$.

As an appropriate geometry for the investigation a parallelepiped is chosen. The corners of the parallelepiped ensure that the condition for the validity of the Rayleigh hypothesis is violated for any source intercepting the minimal sphere. Consequently, the minimal sphere $B_1=B$ encloses the vibrating body completely. For the sake of generality, all lengths and dimensions refer to the x dimension of the source. The side lengths are given as $l_x=1$, $l_y=l_x/3$, and $l_z=2l_x/3$. For simplicity, this parallelepiped has a simple ve-
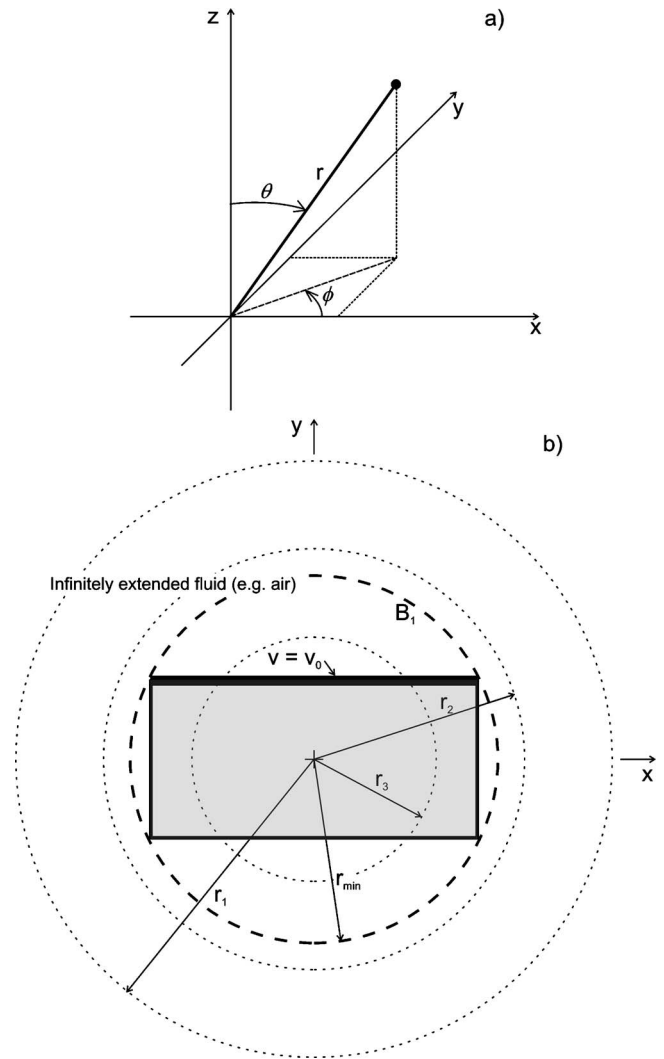
FIG. 1. (a) Spherical coordinate system. (b) Virtual test rig at a section in the equatorial plane. The minimal sphere with radius r_{min} is drawn dashed.

locity distribution at one side and zero velocity at the other. A sinusoidal velocity distribution is applied on a source surface parallel to the x-z-plane generated by

$$v(x,z) = v_0 \cos\left(\frac{\pi x}{l_x}\right)\cos\left(\frac{\pi z}{l_z}\right), \quad (2)$$

with v_0 as velocity amplitude. This distribution suppresses velocity discontinuities at the edges. All walls of the source are assumed rigid except that one vibrating. A sketch of the test rig is shown in Fig. 1 as a section at the equatorial plane. If the KHI given in Eq. (1) is to be used for calculating the sound field radiated from the test source, the velocity as well as the sound pressure on the surface of the source are required.

The pressure can be calculated by employing a discretized form of the KHI. Accordingly, the source surface has to be decomposed into N surface elements S_j. The pressure at the point $\vec{x}_0$, where $\vec{x}_0$ is a point on the source surface, can be approximated by summing over all discrete surface elements as

$$p_N(\vec{x}_0) = \sum_{j=1}^{N} A_j(\vec{x}_0)v_{n,j} + \sum_{j=1}^{N} B_j(\vec{x}_0)p_j \qquad (3)$$

with

$$A_j(\vec{x}_0) = \frac{i\omega\rho_0}{2\pi} \int_{S_j} \frac{e^{-ikr}}{r} dS_j \qquad (4)$$

and

$$B_j(\vec{x}_0) = \frac{1}{2\pi} \int_{S_j} \frac{\partial}{\partial n}\left(\frac{e^{-ikr}}{r}\right) dS_j. \qquad (5)$$

This leads to a set of linear equations given by

$$(\mathbf{I} - \mathbf{B})\mathbf{p} = \mathbf{A}\mathbf{v}, \qquad (6)$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{A}$, $\mathbf{B}$, $\mathbf{p}$, and $\mathbf{v}$ are of the form

$$\mathbf{A} = \begin{bmatrix} A_1(x_1) & A_2(x_1) & \cdots & A_N(x_1) \\ A_1(x_2) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ A_1(x_N) & \cdots & \cdots & A_N(x_N) \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} B_1(x_1) & B_2(x_1) & \cdots & B_N(x_1) \\ B_1(x_2) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ B_1(x_N) & \cdots & \cdots & B_N(x_N) \end{bmatrix},$$

$$\mathbf{p} = \begin{bmatrix} p_1 \\ \vdots \\ \vdots \\ \vdots \\ p_N \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ \vdots \\ \vdots \\ v_N \end{bmatrix}. \qquad (7)$$

The given KHI formulations, however, do not consider the influence of the nondifferentiable source surface, which deteriorates the accuracy of the results. For the exact determination of the radiated sound field the simple discretization chosen should be improved by using triangular or quadrilateral surface elements. The comparatively simple formulation presented in Eqs. (1)–(7), nevertheless, provides accurate results enough for the present purpose. The detailed modeling of the source geometry is of subordinate relevance for comparing the radiated and the reconstructed field at different distances from the source surface.

By using Eq. (1) in conjunction with the known sound pressure and velocity distributions on the surface of the source, the pressure can be calculated at each field point outside the source. With this, on the one hand, a virtual measurement grid of sound pressure values can be generated. On the other hand, pressure and velocity values at arbitrary points can be used for assessing the validity of the NAH reconstruction. The main advantage of the KHI results is that these are not restricted by a minimal sphere limit like NAH.

III. RECONSTRUCTION OF AN ACOUSTIC FIELD

In the following, the reconstruction procedure for the acoustic field is developed. At first, the formulas describing a

spherical wave field are adopted to obtain the fundamentals of NAH. Consider the sound radiation from an arbitrarily shaped sound source that satisfies the Helmholtz equation,

$$(\nabla^2 + k^2)p = 0, \qquad (8)$$

in an unbounded fluid with sound speed c and density ρ_0. The time dependence $e^{j\omega t}$ is suppressed for brevity and the spherical coordinates are indicated in Fig. 1. By considering only outgoing waves, the sound pressure on a sphere with the radius r can obtained as[1]

$$p(r,\theta,\phi) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{h_n^{(2)}(kr)}{h_n^{(2)}(kr_0)} P_{mn}(r_0) Y_n^m(\theta,\phi) \qquad (9)$$

and the velocity accordingly as

$$v(r,\theta,\phi) = \frac{1}{i\rho_0 c} \sum_{n=0}^{N} \sum_{m=-n}^{n} \frac{h_n^{(2)\prime}(kr)}{h_n^{(2)}(kr_0)} P_{mn}(r_0) Y_n^m(\theta,\phi). \qquad (10)$$

The wave field is expanded in terms of a finite series of spherical harmonics,

$$Y_n^m(\theta,\phi) = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m(\cos\theta)e^{im\phi}, \qquad (11)$$

where $P_n^m(x)$ are the associated Legendre functions extended by an orthonormalization factor. The spherical Hankel functions of second kind $h_n^{(2)}(r)$ represent the outward radial propagation. $h_n^{(2)\prime}(r)$ is its derivative with respect to the argument. The series in Eqs. (9) and (10) are truncated after $(N+1)^2$ expansion terms. When the source strength $P_{mn}(r_0)$ is known, the sound pressure $p(\alpha)$ as well as the particle velocity $v(\alpha)$ with $\alpha \in D'$ can be determined at every point α in the three-dimensional sound field outside the source as well as on any point of the source surface as long as α lies outside the minimal sphere B_1.

For determining the source strength $P_{mn}(r_0)$ the sound field has to be decomposed in its spherical wave components. For a known pressure distribution on a sphere of radius r_0 the procedure can be written as

$$P_{mn}(r_0) = \int p(r_0,\theta,\phi)Y_n^m(\theta,\phi)^* d\Omega, \qquad (12)$$

where $( )^*$ indicates the complex conjugate and $d\Omega = \sin\theta d\theta d\phi$ is an angular element. If the sound pressure around the source is sampled at discrete positions the equation has to be discretized.

Alternatively, the wave components coefficients $P_{mn}(r_0)$ can be determined by solving a system of linear equations. By omitting the radial transform that is already included in the back transformation, Eq. (9) is modified to

$$p(r,\theta,\phi) = \sum_{n=0}^{N} \sum_{m=-n}^{n} P_{mn}(r)Y_n^m(\theta,\phi). \qquad (13)$$

By merging the indices to $i = n^2 + n + m + 1$ and neglecting the r dependence, $P_{mn}(r)$ converts to ψ_i. For the angular position (r,θ,ϕ) substituted by $\vec{\alpha}$ the equation simplifies to

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Jens Prager: Approximate reconstruction of sound fields   2069

TABLE I. First 16 NAH expansion coefficients calculated with either integral method or least-squares method ($N=12$, $kl_x=4$).

n	m	Integral [Eq. (12)]	Least squares [Eq. (16)]
0	0	72.106+1.548i	72.069+1.548i
1	−1	0.315+15.914i	0.315+15.914i
1	0	0.002−0.012i	0
1	1	0.315+15.913i	0.315+15.914i
2	−2	−0.516+0.0248i	0.516+0.0248i
2	−1	−0.001	0
2	0	−0.712−0.038i	−0.637−0.004i
2	1	0.001+0.001i	0
2	2	−0.516+0.024i	−0.516+0.024i
3	−3	0.002+0.073i	0.002+0.073i
3	−2	0	0
3	−1	0.001−0.004i	0−0.003i
3	0	−0.013+0.012i	0
3	1	−0.002−0.002i	0−0.003i
3	2	0	0
3	3	0.002+0.073i	0.002+0.073i

$$p(\vec{\alpha}) = \sum_{i=1}^{I} \psi_i Y_i(\vec{\alpha}) \qquad (14)$$

with $I=(N+1)^2$. For a number of J grid points at $\vec{\alpha}_j$, $j=1,\ldots,J$ the equation can be written in matrix form

$$\begin{bmatrix} p(\vec{\alpha}_1) \\ \vdots \\ p(\vec{\alpha}_J) \end{bmatrix} = \begin{bmatrix} Y_1(\vec{\alpha}_1) & \cdots & Y_1(\vec{\alpha}_J) \\ \vdots & \ddots & \vdots \\ Y_I(\vec{\alpha}_1) & \cdots & Y_I(\vec{\alpha}_J) \end{bmatrix} \cdot \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_I \end{bmatrix} \qquad (15)$$

or short

$$\mathbf{p} = \mathbf{Y} \cdot \mathbf{\psi}. \qquad (16)$$

The wave expansion coefficients $\mathbf{\psi}$ are determined by least-squares fitting Eq. (16) to the data $\mathbf{p}$ determined, for example, from the test setup. The least-squares fit is applicable if the equation is overdetermined or quadratic. Otherwise, Eq. (16) has to be solved by singular value decomposition.

Table I shows the first 16 coefficients $P_{mn}(r_1=7l_x)$ for the test setup calculated with either the integral formula (12) or the least-squares method (16). The configuration of the setup is detailed in Sec. II. By comparing the results it can be seen that both methods deliver essentially the same results for the significant source components. This trend also does not change if the input data are distorted by adding random noise. For the investigations presented, the least-squares method has been employed.

It is worth noticing that the design of the sample grid determines the upper frequency limit of the NAH procedure. Commonly, the sound field is sampled at the crossings of a uniform longitude-latitude grid. The main disadvantage of this design is that the lattice points lie much denser near the poles than in the equatorial region. Especially for a limited number of microphones, such a point distribution is not suitable for NAH application, bearing in mind that the number of sampling points defines an upper limit of the useful frequency range. In Ref. 12 an algorithm is presented to compute optimal facility locations on the unit sphere with high accuracy and an appropriate algorithm to calculate the corresponding weighting factors for application in Eq. (12). The algorithm aims at a minimization of the maximum distance between adjacent grid points. The result of this procedure is a grid of sampling points with a better utilization of the number of microphones available, which increases the upper frequency limit by a factor of 2.[13,14]

FIG. 2. Reconstructed sound pressure ($\cdots$) along a circle at the equator ($\theta=\pi/2$) of radius $r_2>r_{min}$ compared with the KHI (—) result ($kl_x=4$, $r_2=0.66l_x$, $N=12$).

IV. SHRINKING OF THE RECONSTRUCTION SURFACE

The results of the NAH-reconstructed sound field, calculated from KHI-generated virtual measurements and the KHI-sound field results agree as long as $r_2 > \sqrt{(l_x/2)^2+(l_y/2)^2+(l_z/2)^2}$. Fulfilling this condition, the field is reconstructed outside the minimal sphere B_1. Figure 2 shows the results. For the calculation, the pressure values are measured at 144 equally spaced microphone positions on a spherical surface of radius $r_1=7l_x$ and the sound field is reconstructed on radius $r_2=2l_x/3$ (see Fig. 1).

If the condition for the Rayleigh hypothesis is violated by intercepting the test source at its edges with a reconstruction surface of radius $r_3<\sqrt{(l_x/2)^2+(l_y/2)^2+(l_z/2)^2}$ an oscillation effect can be observed. As depicted as the dotted line in Fig. 3 for $r_3=0.46l_x$, the curve of the reconstructed field starts to oscillate around the true value. This oscillation becomes more significant, for smaller radii of the reconstruction surface. As expected, the solution does not converge if there are singularities in the field between measurement surface and reconstruction surface. This is also consistent with Ref. 15. There it is shown that the partial wave expansion as given in Eq. (9) diverges for any point inside B_1. Although it is inherently impossible to obtain accurate local information (point values), as is evident from Fig. 3, the solution contains the features of the global behavior. This corresponds to the propositions made in Ref. 10 regarding scattering. It is

Jens Prager: Approximate reconstruction of sound fields
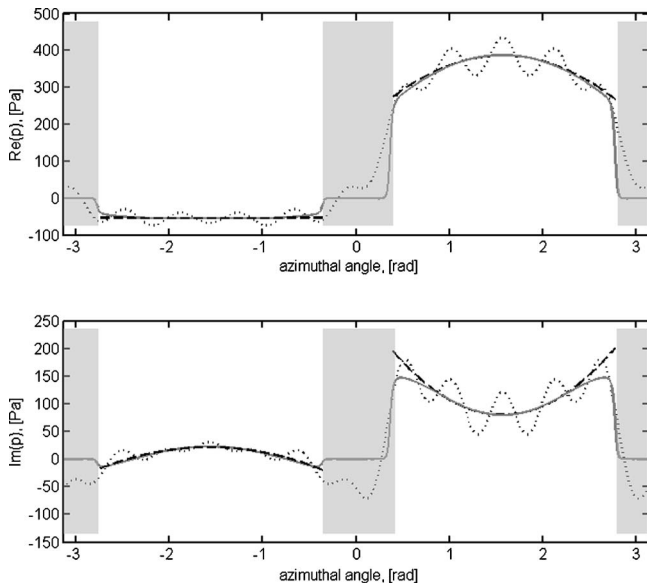
FIG. 3. Comparison of sound pressures reconstructed from NAH with Gegenbauer correction (---) and without (···) with actual values (KHI) (—) on $r_3 < r_{min}$. Intercepted parts are marked gray. ($kl_x = 4$, $r_3 = 0.46l_x$, $N = 12$).

shown that the velocity reconstruction converges at least overall but not pointwise on the source surface for any obstacle shape.

Granted that there is a possibility to extract information about the field also from the nonconvergent description, the solution in the region inside B_1 can be improved at least for limited interceptions. If the nonconvergent series, which in fact contains enough information about the true field, can be converted to a convergent series retaining the same information, the global information can be extracted such that the field can be reconstructed for any point. In its broadest sense the convergence problem described is analogous to Gibbs phenomenon. Although the cause of Gibbs phenomenon is different, the problem resulting in the jump discontinuities is comparable to the NAH singularity problem. Thus it is expected that a resolution of the Gibbs phenomenon can also be applied successfully for the present problem. Such a resolution for problems in spherical coordinates is presented in Refs. 16 and 17. It is based on converting the underlying regular Sturm-Liouville problem to a singular Sturm-Liouville problem which converges rapidly. Assuming that the first $(N+1)^2$ coefficients of the spherical wave expansion can provide enough information to reconstruct the coefficients of an expansion based on a singular Sturm-Liouville problem, the prediction might be improved. This can be done by constructing a Gegenbauer expansion based on the known first few spherical harmonic coefficients. The convergence of the Gegenbauer series has been shown to be exponentially converging on any subinterval for which the function described by the harmonic coefficients is analytic.[18] The latter condition is fulfilled at least if one handles those subareas separately where the intercepting reconstruction surface is outside the source and which do not include singularities. A detailed mathematical proof is given elsewhere[16,17] and will not be part of this paper. In the expansion, the Gegenbauer polynomials[19] $C_n^{(\alpha)}(x)$ for $\alpha \geq 0$ are orthogonal within the

interval $[-1,1]$ with the weighting function $(1-x^2)^{\alpha-1/2}$. The goal is to recover the sound pressure $p(\theta, \phi)$ for ϕ in the subinterval $[\phi_1, \phi_2] \subset [0, 2\pi]$ and a fixed $\theta_{const} \in [0, \pi]$. Adapted to the spherical problem, the Gegenbauer partial sum is defined by

$$\tilde{p}_m^{(\alpha)}(\theta_{const}, \phi) = \sum_{l=0}^{m} \gamma^{(\alpha)}(l) C_l^{(\alpha)}(\xi). \quad (17)$$

For a pressure distribution $p_N(\theta, \phi)$, calculated using Eq. (9) up to the order $N < \infty$, the first m Gegenbauer coefficients are obtained from

$$\gamma^{(\alpha)}(l) = \frac{1}{\Psi_l^\alpha} \int_{-1}^{1} (1-\xi^2)^{\alpha-1/2} C_l^{(\alpha)}(\xi) p_N(\varepsilon\xi + \delta, \theta_{const}) d\xi \quad (18)$$

with the normalization constants

$$\Psi_n^\alpha = \pi^{1/2} C_n^{(\alpha)}(1) \frac{\Gamma(\alpha + 1/2)}{\Gamma(\alpha)(n + \alpha)} \quad (19)$$

and $0 \leq m \leq N$ and $0 \leq \alpha \leq N$. The parameters α and m should be optimized to reduce the error of the procedure. Details about the optimization process are published elsewhere.[16] In the investigations presented the parameters have been set to $m = \alpha = N/4$ according to Ref. 16. Note, that the result of the Gegenbauer expansion $\tilde{p}_m^{(\alpha)}$ is the new sound pressure field.

The angle ϕ in the subinterval $[\phi_1, \phi_2]$ has been transformed to ξ in $[-1, 1]$. The relation between the local variable ξ and the angle ϕ with $\phi_1 < \phi_2$ is defined by

$$\xi(\phi) = \frac{\phi - \delta}{\varepsilon} \quad (20)$$

with

$$\varepsilon = \frac{\phi_2 - \phi_1}{2} \quad (21)$$

and

$$\delta = \frac{\phi_2 + \phi_1}{2}. \quad (22)$$

The same scheme can be applied for θ in lateral direction for constant ϕ. Fully analogous, the velocity field $v_N(\theta, \phi)$ can be treated. The process is repeated for each closed subset of the reconstruction surface outside the source and free of singularities. Consequently, each subset gets its own set of Gegenbauer coefficients. It is worth mentioning that the procedure described is computationally reliable also when N is small.

V. NUMERICAL EXAMPLE

Consider the radiation problem outlined in Sec. II. The task is to reconstruct the acoustic field variables at points close to or on the source surface, starting from measurements of the sound pressure on a spherical measurement grid of radius $r_1 = 7l_x$. For a parallelepiped with a sinusoidal velocity distribution on one x-z surface as described, the radiated

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Jens Prager: Approximate reconstruction of sound fields   2071
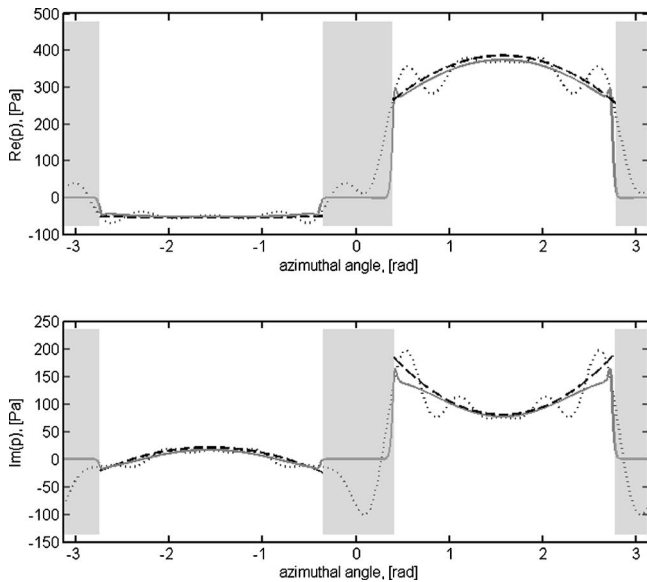
FIG. 4. Comparison of sound pressures reconstructed from NAH with Gegenbauer correction (---) and without ($\cdots$) with actual values (KHI) (—) on $r_3 < r_{\min}$. Intercepted parts are marked gray ($kl_x=4$, $r_3=0.46l_x$, $N=12$, $\theta = \pi/2+0.1$).

sound pressure at the grid is calculated using the KHI. For points on a closed subset of the reconstruction surface inside the minimal sphere but outside the source, the reconstruction has been performed in the following steps:

Step 1: The sound field is expanded in a truncated series of N spherical harmonic wave components. For the results shown in the following, the solution of the system of linear equations as described in Eq. (16) was used. An equally spaced measurement grid as described in Ref. 12 is used.

Step 2: By using Eq. (9) or (10) the sound field variables are reconstructed on a circular arc of radius r and fixed longitude or latitude, respectively. The curve described by the arc must not intercept the source at any point. No singularities are on the curve.

Step 3: The dependent angle on the latitude or longitude interval is transformed to an interval $[-1,1]$ using Eqs. (20)–(22).

Step 4: The $p_N(\theta,\phi)$ or $v_N(\theta,\phi)$ are expanded as truncated series of Gegenbauer polynomials. The resulting series converges exponentially.

The results shown as dashed line in Fig. 3 are calculated for the test rig sketched in Fig. 1. The reconstructed sound pressure is depicted for a circle at the equatorial plane of radius $r_3=0.46l_x$. Due to the interception, the Gegenbauer polynomial expansion is undertaken for two subintervals. To avoid any influence of symmetries of the test rig, in Fig. 4 the results are also presented for a plane at slightly elevated latitude. As can be seen, the results are comparable to those of Fig. 3. The curves show that the additional Gegenbauer polynomial expansion approximation eliminates the oscillation in each subinterval. With a maximum order of spherical wave components of $N=12$, only three Gegenbauer coefficients are used for the approximation. The deviation of the approximation at the ends of each subinterval depends on the number of coefficients used. It is expected that a higher number of coefficients will improve the accuracy of the approxi-
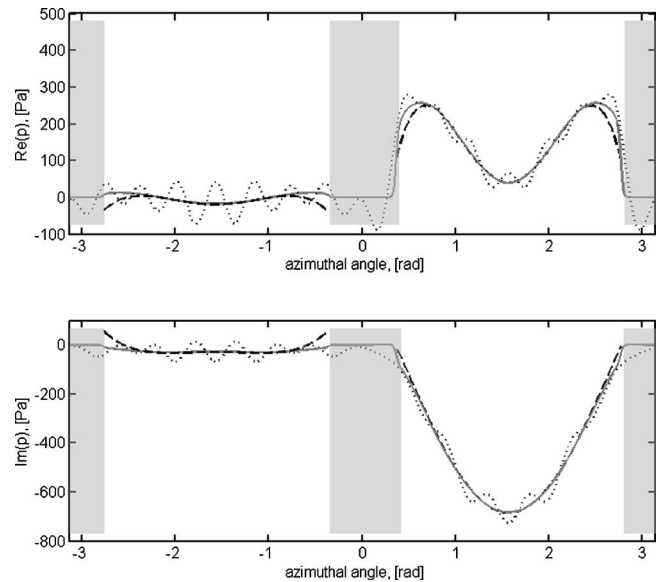
FIG. 5. Comparison of sound pressures reconstructed from NAH with Gegenbauer correction (---) and without ($\cdots$) with actual values (KHI) (—) on $r_3 < r_{\min}$. Results are shown for an enlarged wave number range and a high number of wave components included. Intercepted parts are marked gray ($kl_x=8$, $r_3=0.46l_x$, $N=16$).

mation. This improvement can be seen in Fig. 5 for $N=16$, which allows the use of four Gegenbauer coefficients. Hence, the approximation becomes more and more accurate the higher the number of useful wave components. It is worth noticing that by increasing the frequency as additionally shown in Fig. 5, the oscillation effect diminishes. For the results shown in Fig. 6 the measurement data for a setup as used earlier are distorted by adding of random noise, with a signal to noise ratio of 40 dB. This leads to a mathematically ill posed problem. To get stable results, thus, the maximum order of wave components has to be reduced to $N=9$. Again,
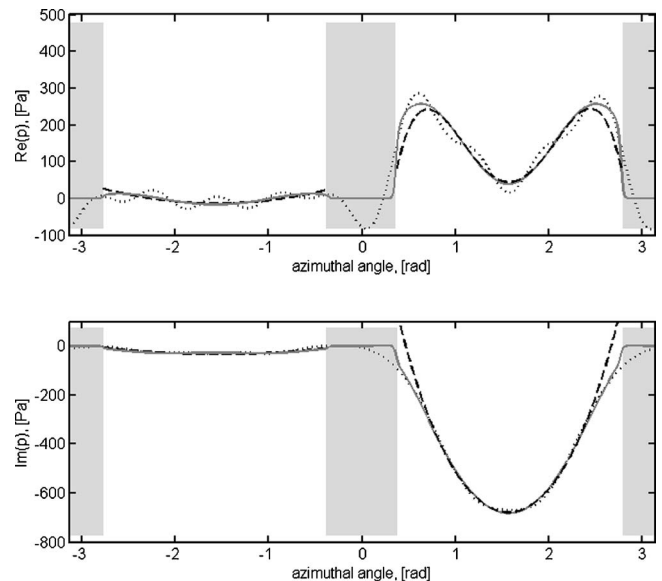
FIG. 6. Comparison of sound pressures reconstructed from NAH with Gegenbauer correction (---) and without ($\cdots$) with actual values (KHI) (—) on $r_3 < r_{\min}$. Noise is added to the input data with a SNR of 40 dB. Intercepted parts are marked gray ($kl_x=8$, $r_3=0.46l_x$, $N=9$).

four Gegenbauer coefficients are used. By considering these constraints it is assumed that the least-squares fit used for determining the expansion coefficients as well as the Gegenbauer expansion are appropriate under practical conditions.

VI. CONCLUDING REMARKS

A method is proposed to overcome the problems arising from singularities in conjunction with NAH sound field reconstruction in spherical coordinates. The occurrence of the singularities depends on the shape of the vibrating body and leads to a nonconvergent solution of the field expansion. An oscillation of the reconstructed sound field variable around the true value arises. It is shown, however, that the incomplete solution contains enough information to recover the sound field approximately. By using a truncated series of Gegenbauer polynomials, the nonconvergent spherical harmonic expansion from the NAH solution can be transformed to a rapidly convergent series. The oscillation is "filtered" and the global information is used for an approximate reconstruction of the field over partial surfaces.

Based on the numerical analysis undertaken, it is argued that the method proposed is limited to low interception ratios. The interception ratio is defined as the ratio of reconstruction surface outside the source to the overall surface of the reconstruction sphere. This means that the method may fail when the interception ratio becomes too small. However, an improved prediction at the limits of each subinterval of the Gegenbauer expansion can be achieved with an increased number of wave components N included. An optimization of N seems to be necessary as well as an investigation of the minimum interception ratio. All these limitations are frequency dependent and depend also on the character of the underlying radiation problem. For engineering practice, one can speculate that the smoother the source shape and the field distribution the less likely is the presence of singularities and the more accurate is the reconstruction as well as the approximation.

[1] M. Ochmann, "The source simulation technique for acoustic radiation problems," Acustica **81**, 512–527 (1995).

[2] Lord Rayleigh, "On the dynamical theory of gratings," Proc. R. Soc. London, Ser. A **79**, 399–416 (1907).

[3] R. F. Millar, "The Rayleigh hypothesis and a related least-square solution to scattering problems for periodic surfaces and other scatterers," Radio Sci. **8**, 785–796 (1973).

[4] P. M. van den Berg and J. T. Fokkema, "The Rayleigh hypothesis in the theory of diffraction by a cylindrical obstacle," IEEE Trans. Antennas Propag. **AP-27**, 577–583 (1979).

[5] D. N. Ghosh Roy and E. G. Williams, "The Rayleigh hypothesis and near-field acoustical holography," Proceedings of the 17th International Congress on Acoustics (ICA), Rome, 2001.

[6] Z. Wang and S. F. Wu, "Helmholtz equation-least-square method for reconstructing the acoustic pressure field," J. Acoust. Soc. Am. **102**, 2020–2032 (1997).

[7] E. G. Williams, *Fourier Acoustics, Sound Radiation and Nearfield Acoustical Holography* (Academic, New York, 1999).

[8] W. A. Veronesi and J. D. Maynard, "Digital holographic reconstruction of source with arbitrarily shaped surfaces," J. Acoust. Soc. Am. **85**, 588–598 (1989).

[9] S. F. Wu, "On reconstruction of acoustic pressure fields using the Helmholtz equation least squares method," J. Acoust. Soc. Am. **107**, 2511–2522 (2000).

[10] A. G. Ramm, "Modified Rayleigh conjecture and applications," J. Phys. A **35**, L357–L361 (2002).

[11] A. D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications* (McGraw-Hill, New York, 1981).

[12] J. Fliege and U. Maier, "A two-stage approach for computing cubature formulae for the sphere," Ergebnisberichte Angewandte Mathematik, No. 139T, Fachbereich Mathematik, Universität Dortmund, 44221 Dortmund, Germany, 1996. An online version is available at ⟨www.mathematik.uni-dortmund.de/lsx/research/projects/fliege/nodes/nodes.html⟩ (last viewed on April 12, 2007).

[13] M. Taylor, "Cubature for the sphere and the discrete spherical harmonic transform," SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **32**, 667–670 (1995).

[14] R. Duraiswami, D. Zotkin, Z. Li, E. Grassi, N. Gumerov, and L. Davis, "System for capturing of high order spatial audio using spherical microphone array and binaural head-tracked playback over headphones with head related transfer function cues," Convention Paper presented at the 119th AES Convention, New York, 2005.

[15] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory* (Springer, Berlin, 1992).

[16] A. Gelb, "The resolution of the Gibbs phenomenon for spherical harmonics," Math. Comput. **66**, 699–717 (1997).

[17] D. Gottlieb and C. Shu, "On the Gibbs phenomenon and its resolution," SIAM Rev. **39**, 644–668 (1997).

[18] D. Gottlieb and C. Shu, "On the Gibbs phenomenon IV. Recovering exponential accuracy in a subinterval from a Gegenbauer partial sum of a piecewise analytic function," Math. Comput. **64**, 1081–1095 (1995).

[19] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).

Prediction of noise changes due to traffic speed control

Rufin Makarewicz[a)] and Piotr Kokowski

Institute of Acoustics, A. Mickiewicz University, 61-614 Poznan, Umultowska 85, Poland

The effects of vehicle speed variation on road traffic noise are analyzed. The steady speed motion is replaced by deceleration, cruise, and acceleration. Because of a relatively loud acceleration noise, such a speed variation results not only in the noise decrease zones, but in the noise increase zones as well. The location of these zones depends slightly upon the ground covering (grass, concrete, etc.). Conversely, their boundaries change dramatically with the parameters describing noise emission during deceleration, cruise, and acceleration. For example, the Japanese and Polish models of noise emission have been applied. The critical length L^* of the cruise segment of the road is introduced: for $L > L^*$ the sound energy decline (due to speed reduction) compensates the sound energy growth (due to acceleration). The results obtained could be useful for road administrators. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2769972]

I. INTRODUCTION

The problem of interrupted traffic noise due to cross-roads with traffic lights[1–10] and roundabouts[11–15] is almost 30 years old. Most recently the noise in the vicinity of road bumps, humps, and speed cushions has been studied.[16–18] Any speed reducer could be seen as a device of the road noise mitigation.[19–22] The main result in this study is the analysis of noise changes due to traffic speed variation (cruise, deceleration, acceleration, cruise). The noise changes are expressed in terms of the A-weighted sound exposure at the receiver (Sec. IV) and by the A-weighted sound energy emitted by a vehicle (Sec. V).

The sound power of noise emitted by a road vehicle declines with its speed. As the speed decrease, $V_1 \rightarrow V_2$, precedes the speed increase, $V_2 \rightarrow V_1$ (Fig. 1), the sound power increase (during acceleration) follows its decrease (during deceleration). Thus, the following question arises: is the temporary reduction of traffic speed an effective control measure for road traffic noise?

We make use of the A-weighted sound exposure of noise which comes from a single vehicle pass by. The approach is based on the experimentally confirmed Japanese emission model, which relates the A-weighted sound power level and the traffic speed in the following way:[23]

$$L_{\mathrm{WA}} = L + 10\mathrm{m} \cdot \lg(V/V_o), \tag{1}$$

where $m=1$ and $m=3$ for the acceleration and deceleration noise, respectively (Sec. III). To get the nondimensional argument of the logarithm, the vehicle speed V is referred to the unit velocity $V_o=1$ (either in km/h or in m/s). Thus the constant L (Eq. (1)) has no physical meaning for it expresses the sound power level at the unit speed $V=V_o$. Owing to the integer value of m, the Japanese emission model makes it possible to calculate quite easily the A-weighted sound exposure, E (Sec. IV) and the A-weighted sound power e (Sec.

V). The emission models, with the formulas similar to Eq. (1) and $m \neq 1,2,3,\ldots$. (see the papers cited above), require numerical methods for E and e computation. In such a case the final results become more complex as compared with the results of this study.

II. NOISE REDUCTION

The A-weighted sound exposure is determined by the integral

$$E = \int_{\tau} p_A^2 dt, \tag{2}$$

where τ is the time interval of noise emission and p_A^2 symbolizes the A-weighted squared sound pressure of noise generated by a single vehicle. In the case under consideration, the road is composed of two cruise segments where the vehicle is moving at a steady speed V_1 (Fig. 1). The relevant A-weighted sound exposures are $E_{s1}(-\infty)$ and $E_{s1}(+\infty)$. The third cruise segment, with the steady speed $V_2 < V_1$, is characterized by E_{s2}. Finally, the values of the A-weighted sound exposures E_d and E_a correspond to the deceleration and acceleration segments of the road. The total noise at the receiver $O(X,D,H)$ (Fig. 2), coming from all road segments, is described by the sum

$$E(+X) = E_{s1}(-\infty) + E_d + E_{s2} + E_a + E_{s1}(+\infty). \tag{3}$$

Without any traffic speed limitation, for a vehicle moving from $x=-\infty$ to $x=+\infty$ at a steady speed V_1, the noise would be characterized by the sound exposure $E_{s1}(\pm\infty)$. Thus, the noise reduction owing to the traffic speed variation ($V_1 \rightarrow V_2 \rightarrow V_1$) can be calculated from

$$\Delta L = 10 \lg \left\{ \frac{E(+X)}{E_{s1}(\pm\infty)} \right\}. \tag{4}$$

The above relationships holds true for a one way traffic, with the vehicles moving from left to right (Fig. 1). A more realistic traffic model is composed of two types of vehicles:

[a)]Author to whom correspondence should be addressed. Electronic mail: makaaku@amu.edu.pl
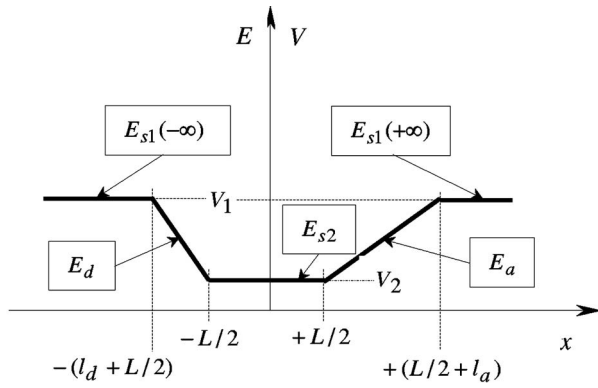
FIG. 1. A road consists of three cruise segments (with the vehicle speeds V_1, V_2, V_1 and the A-weighted sound exposures $E_{s1}(-\infty)$, E_{s2}, $E_{s1}(+\infty)$), deceleration segment (with E_d) and acceleration segment (with E_a).

- moving from $x=-\infty$ to $x=+\infty$ and characterized by $E(+X)$ (Fig. 1),
- moving from $x=+\infty$ to $x=-\infty$ and characterized by $E(-X)$ (Fig. 1 with E_d replaced by E_a and vice versa).

Thus Eq. (4) has to be rearranged into

$$\Delta L = 10 \lg \left\{ \frac{E(+X) + E(-X)}{2E_{s1}(\pm\infty)} \right\}. \qquad (5)$$

The explicit forms of the functions $E(+X)$, $E(-X)$, and $E_{s1}(\pm\infty)$ will be derived in due course. Actually, there are a few categories of road vehicles. To show the salient features of the method, only automobiles are considered in this study.

III. A-WEIGHTED SQUARED SOUND PRESSURE

The results presented in Refs. 15 and 24 show that the nondirectional monopole could be used as a feasible model of a real automobile. We assume that the receiver height exceeds the source height, $H \gg h$ (Fig. 2). Consequently, the distance between the source $S(x,0,h)$ and the receiver $O(X,D,H)$ can be approximated by

$$d \approx \sqrt{(x-X)^2 + D^2 + H^2}. \qquad (6)$$

In an open space, without any obstacles such as buildings, the A-weighted squared sound pressure equals

$$p_A^2 = \frac{W_A(V)\rho c}{4\pi d^2} G(d), \qquad (7)$$

where ρc expresses the characteristics impedance of air, the A-weighted sound power $W_A(V)$ depends on the vehicle speed V, and the function $G(d)$ describes the ground effect. With $H \gg h$, one can write
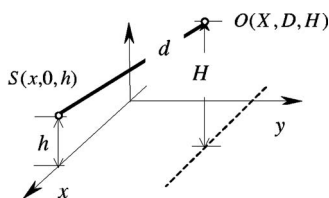
FIG. 2. Distance d (Eq. 6) between the vehicle S and the receiver O.

$$G(d) \approx \frac{\beta}{1 + \frac{1}{\zeta}\left(\frac{d}{H}\right)^2}, \qquad (8)$$

where β and ζ characterize the sound reflection from the road surface and from the adjacent ground, respectively.[25–27] The latter is closely correlated with the ground impedance. To separate the geometrical spreading and the ground effect, we rewrite the above equation as follows:[28]

$$p_A^2 = \frac{\beta W_A(V)\rho c}{4\pi} \left[\frac{1}{d^2} - \frac{1}{d^2 + \zeta H^2} \right]. \qquad (9)$$

Above a hard surface, characterized by a high value of the ground parameter ζ, and the receiver so close to the source that $d \ll \sqrt{\zeta} \cdot H$, sound wave propagation is solely governed by geometrical spreading (Eq. (9))

$$p_A^2 = \frac{\beta W_A(V)\rho c}{4\pi d^2}, \qquad (10)$$

where the product βW_A denotes the modified A-weighted sound power.

The simultaneous measurements of the A-weighted sound pressure level

$$L_A = 10 \lg\{p_A^2/p_o^2\}, \quad p_o = 20\mu\text{Pa}, \qquad (11)$$

by two microphones, (d_1, H_1) and (d_2, H_2), gives the value of ground parameter (Fig. 2, Eqs. (7), (8), and (11)),

$$\zeta = \frac{\left(\frac{d_2}{H_2}\right)^2 - q \cdot \left(\frac{d_1}{H_1}\right)^2}{q - 1}, \qquad (12)$$

where

$$q = \left(\frac{d_1}{d_2}\right)^2 10^{0.1[L_{A1} - L_{A2}]}. \qquad (13)$$

A. Example

Let us take the microphone heights $H_1 = H_2 = 1$ m and the distances $d_1 = 15$, $d_2 = 30$ m. For the difference $L_{A1} - L_{A2} = 6.1$ dB (concrete pavement) and $L_{A1} - L_{A2} = 8.0$ dB (grassy area), one gets $\zeta = 36300$ (hard-type ground) and $\zeta = 944$ (soft-type ground).

Measurements of automobile noise in Japan (Ref. 23) indicate that the A-weighted sound power (at a steady speed V and during deceleration with the instantaneous speed V) is given by (Eq. (1))

$$L_{WA}^{(sd)} = L_{sd} + 30 \lg(V/V_o), \qquad (14)$$

where $L_{sd} = 46.7$ dB. Measurements of noise produced by automobiles during acceleration yield (Ref. 23)

$$L_{WA}^{(a)} = L_a + 10 \lg(V/V_o), \qquad (15)$$

where $L_a = 82.3$ dB. Thousands of noise measurements support both Eq. (14) and Eq. (15). The values of L_{sd} and L_a include the effects of tires and pavements.

For automobiles on Japanese roads their difference equals: $L_a - L_{sd} = 35.6$. The preliminary measurements in Po-

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

R. Makarewicz and P. Kokowski: Prediction of noise changes   2075

land yield the relationships similar to Eqs. (14) and (15) with one exception: $L_a - L_{sd} = 29.3$. In general, the cruise, deceleration, and acceleration noise emission in Japan and Poland can be described as follows (Eqs. (1), (14), and (15)):

$$L_{WA}^{(Jap)} = L_{Jap} + 10 \text{ m} \cdot \lg(V/V_o),$$

$$L_{WA}^{(Pol)} = L_{Pol} + 10 \text{ m} \cdot \lg(V/V_o), \tag{16}$$

where $m=3$ and $m=1$. It seems that the emission models in other countries, within the definite range of traffic speed, $V_1 < V < V_2$, could be fitted to Eq. (1).

Making use of the definition of the A-weighted sound power level, one obtains the A-weighted sound power for both steady speed (or deceleration) and the A-weighted sound power for acceleration

$$W_A^{(sd)} = W_{sd} \cdot (V/V_o)^3, \quad W_A^{(a)} = W_a \cdot (V/V_o). \tag{17}$$

Here

$$W_{sd} = W_o \cdot 10^{0.1 L_{sd}}, \quad W_a = W_o \cdot 10^{0.1 L_a}, \tag{18}$$

where the reference sound power $W_o = 10^{-12}$ [Watts] and the values of L_{sd}, L_a are supposed to be known from L_{WA} measurements (Eqs. (14) and (15)).

IV. A-WEIGHTED SOUND EXPOSURE

At a constant deceleration rate a_d, the deceleration segment length is (Fig. 1)

$$\ell_d = \frac{1}{2a_d}(V_1^2 - V_2^2). \tag{19}$$

Similarly, the acceleration segment length equals,

$$\ell_a = \frac{1}{2a_a}(V_1^2 - V_2^2), \tag{20}$$

where a_a is the constant acceleration rate.

As is shown in Fig. 1, a vehicle moves at a steady speed V_1 along the two road segments, $[-\infty < x < -L/2 - l_d]$ and $[+L/2 + l_a < x < +\infty]$. The reduced traffic speed V_2 is ascribed to the middle road segment $[-L/2 \leq x \leq +L/2]$. Deceleration and acceleration take place on the segments $[-L/2 - l_d < x < -L/2]$ and $[+L/2 < x < L/2 + l_a]$, respectively.

A. Steady speed noise

For any steady speed V, the infinitesimal displacement of a vehicle equals $dx = Vdt$, and the A-weighted sound exposure (Eq. (2)) can be rewritten as

$$E = \int_\ell \frac{p_A^2}{V} dx, \tag{21}$$

where ℓ is length of the road segment. Accordingly, the A-weighted sound exposure for the steady speed motion can be expressed as follows (Eqs. (9), (17), (18), and (21)):

$$E_s = \frac{\rho c}{4\pi} \int_\ell S_{sd} \left[\frac{1}{d^2} - \frac{1}{d^2 + \zeta H^2} \right] dx. \tag{22}$$

Here noise emission quantifies the function of the vehicle speed

$$S_{sd} = \frac{\beta W_o}{V_o} \cdot 10^{0.1 L_{sd}} \cdot \left(\frac{V}{V_o} \right)^2. \tag{23}$$

Dimensional analysis shows that S_{sd} is related to the sound energy emitted from the line source of unit length (Joules per meter). Ultimately, Eqs. (6), (22), and (23) yield the A-weighted sound exposure of noise radiated from the road segment of length ℓ

$$E_s = \frac{\beta W_o \rho c}{4\pi D_o} 10^{0.1 L_{sd}} \frac{V^2}{V_o^3} [F_s(0) - F_s(\zeta)], \tag{24}$$

where $D_o = 1$ m and

$$F_s(\zeta) = D_o \int_\ell \frac{dx}{(x - X)^2 + D^2 + (1 + \zeta)H^2}. \tag{25}$$

The noise from both cruise segments $[-\infty < x < -L/2 - l_d]$ and $[+L/2 + l_a < x < +\infty]$, with the steady speed V_1 (Fig. 1), is characterized by

$$E_{s1}(+\infty) + E_{s1}(-\infty) = \frac{\beta W_o \rho c}{4\pi D_o} 10^{0.1 L_{sd}} \frac{V_1^2}{V_o^3} [F_{s1}(0) - F_{s1}(\zeta)], \tag{26}$$

where

$$F_{s1}(\zeta) = \frac{D_o}{r} \left[\pi - \arctan \frac{L/2 + \ell_d + X}{r(s)} - \arctan \frac{L/2 + \ell_a - X}{r(s)} \right], \tag{27}$$

and

$$r^2(\zeta) = D^2 + (1 + \zeta)H^2. \tag{28}$$

Let us calculate the A-weighted sound exposure, $E_{s1}(\pm\infty)$ (Eq. (4)), of noise emitted by the vehicle cruising (at a steady speed V_1) from $x = -\infty$ to $x = +\infty$, without any interruptions. Setting $L=0$, $\ell_d=0$ and $\ell_a=0$, one arrives at (Eqs. (26) and (27)),

$$E_{s1}(\pm\infty) = \frac{\beta W_o \rho c}{4} 10^{0.1 L_{sd}} \frac{V_1^2}{V_o^3} \left[\frac{1}{r(0)} - \frac{1}{r(\zeta)} \right], \tag{29}$$

where $r(\zeta)$ is defined by Eq. (28). The middle road segment $[-L/2 \leq x \leq +L/2]$ with the reduced speed V_2 (Fig. 1) yields the noise which is characterized by (Eqs. (24) and (25))

$$E_{s2} = \frac{\beta W_o \rho c}{4\pi D_o} 10^{0.1 L_{sd}} \frac{V_2^2}{V_o^3} [F_{s2}(0) - F_{s2}(\zeta)], \tag{30}$$

where

$$F_{s2}(\zeta) = \frac{D_o}{r(\zeta)} \left[\arctan \frac{L/2 - X}{r(\zeta)} + \arctan \frac{L/2 + X}{r(\zeta)} \right]. \tag{31}$$

B. Deceleration noise

For a constant deceleration rate a_d, the infinitesimal time increment equals $dt=-dV/a_d$, and the A-weighted sound exposure (Eq. (2)) takes the form (Fig. 1)

$$E_d = -\frac{1}{a_d}\int_{V_1}^{V_2} p_A^2 dV. \tag{32}$$

Equation (9) shows that the A-weighted squared sound pressure p_A^2 depends on the distance d between the vehicle S and the receiver O (Fig. 1). However, for the time of deceleration the location of the vehicle on the x axis is a function of the instantaneous velocity

$$x(V) = \frac{1}{2a_d}(V_1^2 - V^2) - \left(\frac{L}{2} + \ell_d\right), \tag{33}$$

where $V_2 \leqslant V \leqslant V_1$ (Fig. 1). Thus the distance d (Eq. (6)) becomes a function of the speed V. Accordingly, Eqs. (6) and (33) imply that

$$d^2(V) + \zeta H^2 = \frac{1}{(2a_d)^2}[(V^2 + A_d)^2 + B_d^2], \tag{34}$$

where

$$A_d = (L + 2X)a_d - V_2^2 \quad \text{and} \quad B_d = 2a_d\sqrt{D^2 + (1+\zeta)H^2}. \tag{35}$$

Finally, the A-weighted sound exposure of deceleration noise takes the form (Eqs. (9), (32), and (34))

$$E_d = \frac{\beta W_o \rho c}{2\pi}10^{0.1L_{sd}}\frac{a_d}{V_o^3}[F_d(0) - F_d(\zeta)], \tag{36}$$

where

$$F_d(\zeta) = \int_{V_2}^{V_1}\frac{2V^3 dV}{(V^2 + A_d)^2 + B_d^2}. \tag{37}$$

The above integral can be solved,

$$F_d = \ln\sqrt{\frac{(X + L/2 + \ell_d)^2 + r^2}{(X + L/2)^2 + r^2}} + \frac{X + L/2 - V_2^2/2a_d}{r}$$
$$\times\left[\arctan\frac{X + L/2}{r} - \arctan\frac{X + L/2 + \ell_d}{r}\right], \tag{38}$$

where the deceleration length ℓ_d and the distance r and are determined by Eqs. (19) and (28), respectively.

C. Acceleration noise

Similarly as for deceleration, a constant acceleration rate a_a gives the infinitesimal time increment, $dt=dV/a_a$. Accordingly, the A-weighted sound exposure (Eq. (2)) can be written as (Fig. 1)

$$E_a = \frac{1}{a_a}\int_{V_2}^{V_1} p_A^2 dV. \tag{39}$$

Consequently the x-axis coordinate of the accelerating vehicle can be calculated from

$$x(V) = \frac{1}{2a_a}(V^2 - V_2^2) + \frac{L}{2}, \tag{40}$$

where $V_2 \leqslant V \leqslant V_1$. If one substitutes Eq. (40) into Eq. (6) one finds

$$d^2(V) + \zeta H^2 = \frac{1}{(2a_a)^2}[(V^2 + A_a)^2 + B_a^2], \tag{41}$$

where

$$A_a = (L - 2X)a_a - V_2^2 \quad \text{and} \quad B_a = 2a_a\sqrt{D^2 + (1+\zeta)H^2}. \tag{42}$$

Thus, the A-weighted sound exposure of acceleration noise is determined by (Eqs. (9), (39), and (41)),

$$E_a = \frac{\beta W_o \rho c}{4\pi D_o V_o}10^{0.1L_a}[F_a(0) - F_a(\zeta)], \tag{43}$$

where

$$F_a(\zeta) = 2a_a D_o\int_{V_2}^{V_1}\frac{2VdV}{(V^2 + A_a)^2 + B_a^2}. \tag{44}$$

The above integration brings about

$$F_a = \frac{D_o}{r}\left[\arctan\frac{X - L/2}{r} - \arctan\frac{X - L/2 - \ell_a}{r}\right], \tag{45}$$

where the acceleration length ℓ_a and the distance r are defined by Eqs. (20) and (28), respectively.

D. Noise reduction

Using Eqs. (3), (26), (30), (36), and (43) it then follows that the A-weighted sound exposure of noise from a road with the speed variation, $V_1 \to V_2 \to V_1$ (Fig. 1), can be calculated from

$$E(+X) = \frac{\beta W_o \rho c}{4\pi D_o V_o}10^{0.1L_{sd}}\left\{\left(\frac{V_1}{V_o}\right)^2[F_{s1}(0) - F_{s1}(\zeta)]\right.$$
$$+ \left(\frac{V_2}{V_o}\right)^2[F_{s2}(0) - F_{s2}(\zeta)] + \frac{2a_d D_o}{V_o^2}[F_d(0)$$
$$\left. - F_d(\zeta)] + 10^{0.1(L_a - L_{sd})}[F_a(0) - F_a(\zeta)]\right\}, \tag{46}$$

where $D_o=1$ m, $V_o=1$ m/s, and the *emission characteristic* $L_a - L_{sd}=35.6$ for Japan (or $L_a - L_{sd}=29.3$ for Poland). All four functions $F_{s1}(\zeta)=F_{s1}(X,\ldots,\zeta)$ (Eq. (27)), $F_{s2}(\zeta)=F_{s2}(X,\ldots,\zeta)$ (Eq. (31)), $F_d(\zeta)=F_d(X,\ldots,\zeta)$ (Eq. (38)), and $F_a(\zeta)=F_a(X,\ldots,\zeta)$ (Eq. (45)) depend on the X coordinate of the receiver $O(X,D,H)$ (Fig. 2). Thus, changing $X \to -X$ in Eq. (46) one arrives at the explicit form of $E(-X)$ (see Eq. (5)), which corresponds to a vehicle moving from $x=+\infty$ to $x=-\infty$.

Suppose that the length L of the road segment with the reduced speed V_2 (Fig. 1) is much longer than the deceleration length ℓ_d (Eq. (19)) and the acceleration length ℓ_a (Eq. (20)). Then, setting $L=\infty$ and noting Eqs. (5), (29), and (46), one arrives at the simplest formula for the noise reduction,

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

R. Makarewicz and P. Kokowski: Prediction of noise changes   2077

	V_1						
V_2	20	30	40	50	60	70	80
0	1.37	1.50	1.61	1.70	1.78	1.85	1.91
30			1.17	1.33	1.46	1.56	1.63
60						1.13	1.28

$$\Delta L \approx 20 \lg \left\{ \frac{V_2}{V_1} \right\}. \tag{47}$$

The deceleration rate a_d and acceleration rate a_a are functions of the deceleration length (Eq. (19)), acceleration length (Eq. (20)), as well the initial speed V_1 and the reduced speed V_2 (Tables I and II, see Ref. 29). For example, taking $V_1 = 19.5$ [m/s] (i.e., 70 [km/h]) and $V_2 = 11.1$ [m/s] (i.e., 40 [km/h] one can find $a_d = 1.42$ [m/s^2] and $a_a = 1.26$ [m/s^2]. In such a case Eqs. (19) and (20) yield the deceleration length, $\ell_d = 90.5$ m, and the acceleration length, $\ell_a = 102$ m.

E. Example

The above values of a_d, a_d, ℓ_d, ℓ_d, and V_1, V_2 (expressed in m/s) have been used for the calculation of noise reduction $\Delta L(X, Y)$ (Eqs. (5), (29), and (46)). For the emission characteristic $L_a - L_{sd} = 35.6$ dB (describing automobile noise in Japan—see Sec. III), the length of the road segment $L = 200$ m, and the receiver height $H = 4$ m (Fig. 2), the calculations have been made for hard ground ($\zeta = 36\,300$) and soft ground ($\zeta = 944$), respectively. Figures 3–6 show the noise increase zones (where $0 < \Delta L < +2$ dB, $+2 < \Delta L < +4$ dB, ...) and the noise decrease zones (where $0 < \Delta L < -2$ dB, $-2 < \Delta L < -4$ dB, etc.). The influence of the ground effect on the zones' location is rather small. Dramatic change of zones location occurs when the emission characteristic $L_a - L_{sd} = 35.6$ (Japan) is replaced by $L_a - L_{sd} = 29.3$ (Poland). The location of noise decrease and increase zones in Figs. 5 and 6 are completely different from those in Figs. 3 and 4, respectively. Figures 7 and 8 show the results of ΔL calculation for the intermediate values of $L_a - L_{sa} = 31.0$ and $L_a - L_{sa} = 33.0$, respectively.

Anyway, it is not easy to answer the following question: is the reduced speed V_2 sufficiently small and the corresponding length L sufficiently long (Fig. 1), to improve the acoustical quality of the adjacent terrain. To answer this question, the A-weighted sound energy has to be estimated.

	V_1						
V_2	20	30	40	50	60	70	80
0	1.30	1.37	1.44	1.50	1.54	1.57	1.61
30			1.15	1.23	1.30	1.35	1.39
60						1.09	1.17
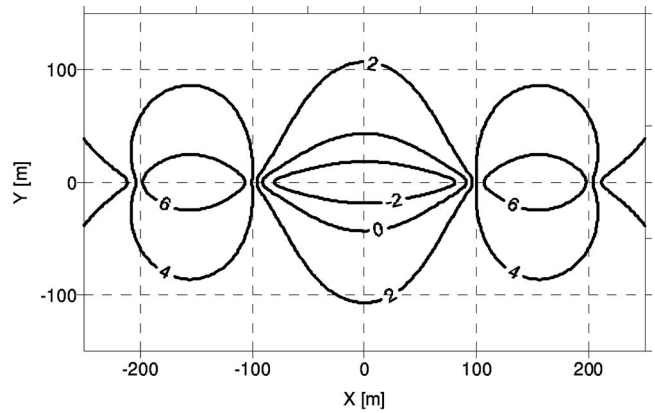
FIG. 3. Noise reduction ΔL at the height $H = 4$ m, calculated for the *hard* ground ($\zeta = 36\,300$) in *Japan* ($L_a - L_{sd} = 35.6$), with the traffic speeds $V_1 = 70$ [km/h] and $V_2 = 40$ [km/h], where the road segment length $L = 200$ m.

V. A-WEIGHTED SOUND ENERGY

Let us consider a road segment with the length, $\ell_d + L + \ell_a$, whose location is defined by inequalities, $-(\ell_d + L/2) < x < +(\ell_a + L/2)$ (see Fig. 1). When the traffic speed changes ($V_1 \rightarrow V_2 \rightarrow V_1$), the A-weighted sound energy radiated from the above defined segment equals

$$e_d + e_s(V_2) + e_a, \tag{48}$$

where e_d denotes the A-weighted sound energy emitted from a vehicle during its deceleration, $e_s(V_2)$ is the A-weighted sound energy for the time interval of vehicle cruising at a reduced speed V_2, and finally e_a expresses the A-weighted sound energy of acceleration. Without any speed variation, the vehicle would move at a constant speed V_1 along the road segment of the length, $\ell_d + L + \ell_a$. Let $e_s(V_1)$ denote the relevant A-weighted sound energy. The traffic speed changes (Fig. 1) improve the acoustical quality of the adjacent terrain, when

$$e_d + e_s(V_2) + e_a < e_s(V_1). \tag{49}$$

In our model the traffic speed V_1 (Fig. 1) and the A-weighted sound energy e, emitted from the road segments $(-\infty, -\ell_d - L/2)$ and $(+L/2 + \ell_a, +\infty)$, remain unchanged. Therefore, these two segments are not considered in this section.
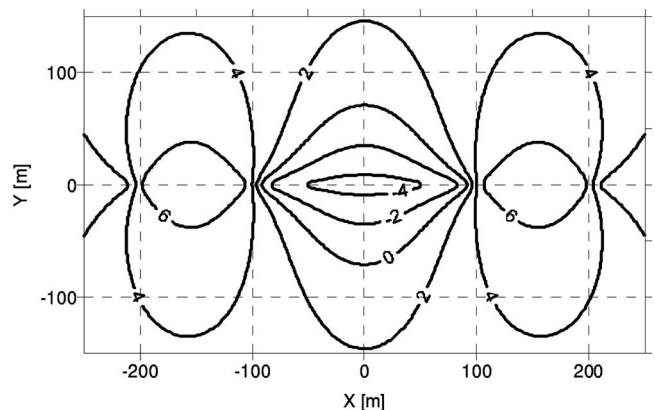
FIG. 4. Noise reduction ΔL at the height $H = 4$ m, calculated for the *soft* ground ($\zeta = 944$) in *Japan* ($L_a - L_{sd} = 35.6$), with the traffic speeds $V_1 = 70$ [km/h] and $V_2 = 40$ [km/h], where the road segment length $L = 200$ m.
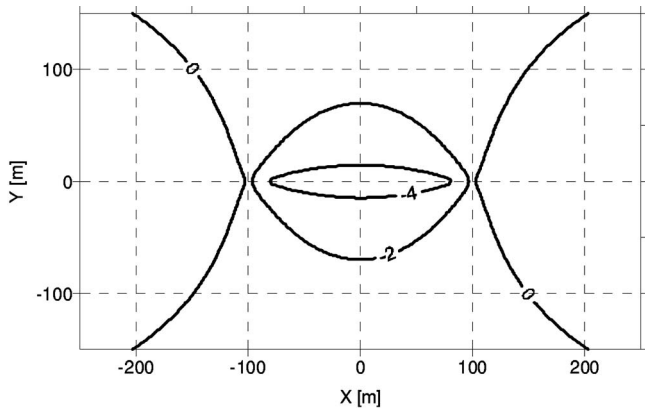
FIG. 5. Noise reduction ΔL at the height $H=4$ m, calculated for the *hard* ground ($\zeta=36\,300$) in *Poland* ($L_a-L_{sd}=29.3$), with the traffic speeds $V_1 =70$ [km/h] and $V_2=40$ [km/h], where the road segment length $L=200$ m.

In general the A-weighted sound energy is determined by

$$e = \int_\tau W_A dt, \tag{50}$$

where τ is the time interval of noise radiation. For the deceleration ($dV=-a_d dt$), acceleration ($dV=a_a dt$), and the steady motion ($\tau=L/V_2$), one gets

$$e_d = \frac{1}{a_d} \int_{V_2}^{V_1} W_A^{(sd)}(V) dV,$$

$$e_a = \frac{1}{a_a} \int_{V_2}^{V_1} W_A^{(a)}(V) dV, \tag{51}$$

$$e_s(V_2) = W_A^{(sd)}(V_2) \frac{L}{V_2},$$

respectively. Taking into account Eqs. (17), (18), and (51), the above relationships yield

$$e_d = \beta W_o 10^{0.1 L_{sd}} \frac{1}{4 a_d V_o^3} [V_1^4 - V_2^4], \tag{52}$$
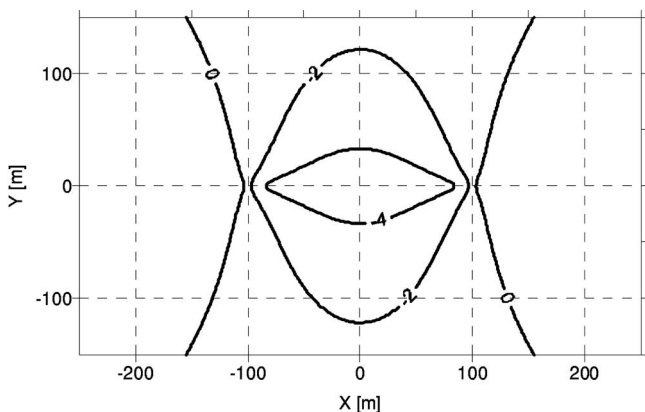
FIG. 6. Noise reduction ΔL at the height $H=4$ m, calculated for the *soft* ground ($\zeta=944$) in *Poland* ($L_a-L_{sd}=29.3$), with traffic speeds $V_1 =70$ [km/h] and $V_2=40$ [km/h], where the road segment length $L=200$ m.
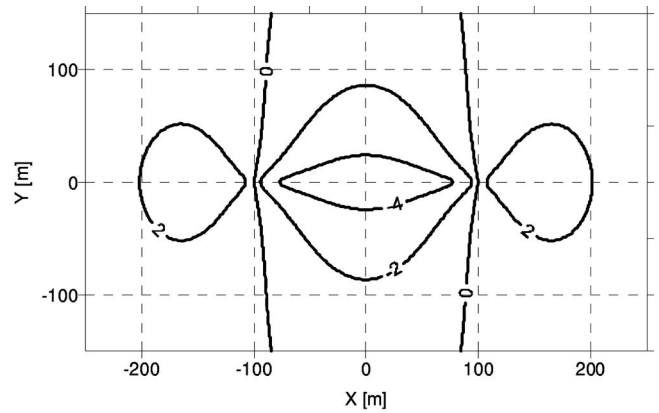
FIG. 7. Noise reduction ΔL at the height $H=4$ m, calculated for the *soft* ground ($\zeta=944$) with the emission characteristic $L_a-L_{sd}=31.0$, traffic speeds $V_1=70$ [km/h] and $V_2=40$ [km/h], where the road segment length $L=200$ m.

$$e_a = \beta W_o 10^{0.1 L_a} \frac{1}{2 a_a V_o} [V_1^2 - V_2^2] \tag{53}$$

and

$$e_s(V_2) = \beta W_o 10^{0.1 L_{sd}} \frac{V_2^2}{V_o^3} L. \tag{54}$$

Without any speed change, the energy radiated from the road segment of length $\ell_d+L+\ell_a$, while the vehicle moves at a steady speed V_1, is given by

$$e_s(V_1) = W_A^{(s,d)}(V_1) \frac{\ell_d+L+\ell_a}{V_1}. \tag{55}$$

Taking into account the A-weighted sound power $W_A^{(sd)}(V)$ (Eqs. (17) and (18)), deceleration length ℓ_d (Eq. (19)), and acceleration length ℓ_a (Eq. (20)), Eq. (55) can be rearranged into the form

$$e_s(V_1) = \beta W_o 10^{0.1 L_{sd}} \frac{V_1^2}{V_o^3} \left[L + (V_1^2 - V_2^2) \frac{a_d+a_a}{2 a_d a_a} \right]. \tag{56}$$

The inequality (49) is satisfied when the length of the road segment L (Fig. 1) exceeds the critical length
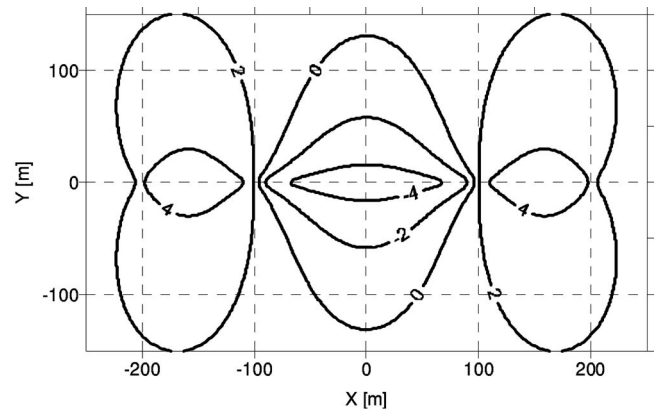
FIG. 8. Noise reduction ΔL at the height $H=4$ m, calculated for the *soft* ground ($\zeta=944$) with the emission characteristics, $L_a-L_{sd}=33.0$, and steady speeds $V_1=70$ [km/h], $V_2=40$ [km/h], where the road segment length $L=200$ m.

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

R. Makarewicz and P. Kokowski: Prediction of noise changes   2079

TABLE III. Critical lengths L_{jap}^* and L_{pol}^* of the road segment (with the reduced speed V_2) calculated from Eq. (58) and Tables I and II), for automobiles driving in Japan ($L_a - L_{\text{sd}} = 35.6$) and in Poland ($L_a - L_{\text{sd}} = 29.3$).

V_1 [km/h]	50	60	70	80
V_2 [km/h]	30	30	30	30
L_{jap}^* [m]	1374	1254	1155	1062
L_{pol}^* [m]	244	185	126	86

$$L > L^*,\tag{57}$$

where L^* is solution to the equation

$$e_d + e_s(V_2, L^*) + e_a = e_s(V_1, L^*).\tag{58}$$

After some calculation and application of Eqs. (51)–(54) and Eq. (56), one arrives at

$$L^* = \frac{V_1^2 + V_2^2}{4a_d} + \frac{V_o^2}{2a_a} 10^{0.1(L_a - L_{\text{sd}})} - V_1^2 \frac{a_d + a_a}{2a_a a_d},\tag{59}$$

where $V_o = 1$ m/s and the speeds V_1, V_2 are expressed in m/s. Making use of Tables I and II, with the emission characteristics $L_a - L_{\text{sd}} = 35.6$ (Japan) and $L_a - L_{\text{sd}} = 29.3$ (Poland), one gets the critical length L_{jap}^* and L_{pol}^*, respectively (Table III). If the actual length of the road segment L exceeds the critical length L^* (Eq. (57)), then the total sound energy $e_d + e_s(V_2) + e_a$ (Eq. (48)) is less than the sound energy $e_s(V_1)$ (see Eq. (56)). The latter is emitted during the undisturbed motion at a steady speed V_1, along the road segment of the length $\ell_d + L + \ell_a$. Figure 9 shows the critical length L^* (Eq. (59)), which is plotted as the function of the emission characteristic, $L_a - L_{\text{sd}}$, for three pairs of the initial and final speeds

$$L^* = \frac{1}{2.44} 10^{0.1(L_a - L_{\text{sd}})} - 143 \text{ [m]},\tag{60}$$

for $V_1 = 60$ [km/h], $V_2 = 40$ [km/h], $a_d = 1.35$ [m/s^2], $a_a = 1.22$ [m/s^2]

$$L^* = \frac{1}{2.52} 10^{0.1(L_a - L_{\text{sd}})} - 196 \text{ [m]},\tag{61}$$

for $V_1 = 70$ [km/h], $V_2 = 40$ [km/h], $a_d = 1.42$ [m/s^2], $a_a = 1.26$ [m/s^2], and finally,
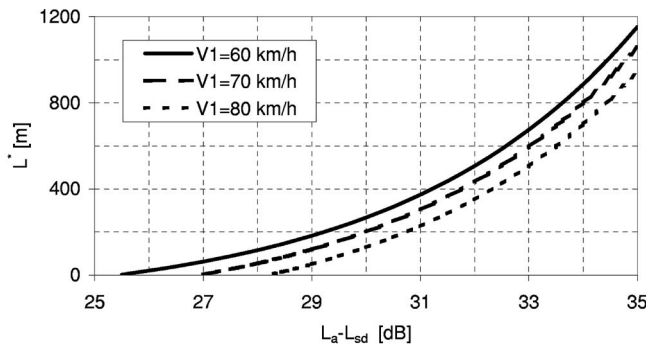
FIG. 9. The critical length L^* as the function of the noise emission characteristic, $L_a - L_{\text{sd}}$, for the traffic speeds ($V_1 = 60$ [km/h], $V_2 = 40$ [km/h]), ($V_1 = 70$ [km/h], $V_2 = 40$ [km/h]), and ($V_1 = 80$ [km/h], $V_2 = 40$ [km/h]). The required deceleration rates, a_d, and acceleration rates, a_d, have been taken from Tables I and II.

$$L^* = \frac{1}{2.64} 10^{0.1(L_a - L_{\text{sd}})} - 250 \text{ [m]},\tag{62}$$

for $V_1 = 80$ [km/h], $V_2 = 40$ [km/h], $a_d = 1.51$ [m/s^2], $a_a = 1.32$ [m/s^2]. The values of a_d and a_a have been found in Tables I and II for each pair of velocities.

VI. CONCLUSIONS

Decrease in the traffic speed from V_1 to V_2 and then its decrease to V_1 again (Fig. 1) gives rise to noise increase zones ($0 < \Delta L < +2$ dB, $+2 < \Delta L < +4$ dB,…) and to noise decrease zones ($0 < \Delta L < -2$ dB, $-2 < \Delta L < -4$ dB,…). Calculation of ΔL from Eqs. (5), (29), and (46) provides their location in the (x, y) plane. To find the adequate length L of the road segment traveled at a reduced speed V_2, the critical length L^* (Eq. (59)) has to be known. The limitations and constrains of the method presented in this study comes from the basic assumptions:

- there is only one category of vehicles, i.e., automobiles,
- an automobile is represented by a nondirectional point source (Eq. (7)),
- deceleration and acceleration rates, a_d and a_a, are constant (Eqs. (19) and (20)),
- the A-weighted sound power W_A, for both cruising and deceleration, is a cubic function of the automobile speed V (Eq. (15)),
- the A-weighted sound power for acceleration is a linear function of the automobile speed (Eq. (16)),
- the homogeneous atmosphere is at rest, so there is no refraction,
- noise propagation is not disturbed by buildings and other obstacles.

Although rather restrictive, these assumptions are quite realistic. It seems that the method presented here captures the salient features of noise emission and propagation. It would become more precise by refined consideration of the above assumption. For example, by including a few categories of vehicles, considering directivity of the point source, and including building reflections and atmospheric refraction. Thus, the answer to the question "how far the presented method is valid" is as follows: as far as the refraction and building's reflections can be neglected.

ACKNOWLEDGMENTS

[1] L. J. M. Jacobs, L. Nijs, and J. J. Van Willingenburg, "A computer model to predict traffic noise in urban situations under free flow and traffic light conditions," J. Sound Vib. **72**, 523–537 (1980).
[2] S. E. Samuels, "The generation of traffic noise under interrupted flow conditions," Ph.D. thesis, Department of Mechanical Engineering, Newcastle University (1988).
[3] P. Kokowski and R. Makarewicz, "Interrupted traffic noise," J. Acoust. Soc. Am. **101**, 360–371 (1997).
[4] W. Bolby, R. L. Wayson, S. Chiguluri, M. Martin, and L. A. Herman, "Interrupted flow reference energy mean emission levels for the FHWA Traffic Noise Model," DOT FHWA-PD-97-019 (1997).
[5] L. Leclercg, J. Lelong, and J. Defrance, "Dynamic assessment of road

traffic noise: Elaboration of a global model," *International Congress on Acoustics 2004*, Kyoto, Japan.

[6] K. Yoshihitsa, Y. Oshino, K. Yamamoto, and H. Tachibana, "Road traffic noise prediction in the vicinity of signalized intersections," *International Congress on Acoustics 2004*, Kyoto, Japan.

[7] H. Jonasson, U. Sandberg, G. van Blokland, J. Ejsmont, G. Watts, and M. Luminari, "Source modeling of road vehicles," HAR11TR-041210-SP10, Harmonoise (2004).

[8] J. A. Ejsmont and G. Ronowski, "Vehicle and tire/road noise during interrupted-flow traffic conditions," *Inter-Noise 2005*, Rio de Janerio.

[9] A. Bhaskar, E. Chung, and A. Dumont, "Study of vehicle noise under different operating," *Sixth Swiss Transport Research Conference*, Monet Verita (2006).

[10] B. De Coensel, D. Botteldooren, F. Vanhoe, and S. Logghe, "Microsimulation based correction on the road traffic noise emission near intersection," Acta. Acust. Acust. **93**, 241–252 (2007).

[11] P. T. Lewis and A. James, "Noise levels in the vicinity of traffic roundabouts," J. Sound Vib. **72**, 51–69 (1980).

[12] W. M. To and T. M. Chan, "The noise emitted from vehicles at roundabouts," J. Acoust. Soc. Am. **107**, 2760–2763 (2000).

[13] M. Berengier, "Acoustical impact of traffic flowing equipment in urban area," *Forum Acusticum 2002*, Sevilla.

[14] M. Bertoni, "Experience in planning of noise mitigation measures in evaluating their effectiveness in the city of Modena," *Inter-Noise 2004*, Prague.

[15] J. Picaut, M. Berengier, and E. Pousseau, "Noise impact modeling of a roundabout," *Inter-Noise 2005*, Rio de Janerio.

[16] P. Bovy, "Carres berlinois et rigoles, evaluation et recommendation d'amengament, Securite dans les rues de quartier in EPFL," Lausanne, Report No. IREC 1007 (1993).

[17] P. Abbott, J. Tyler, and R. Layfield, "Traffic calming-vehicle noise emissions alongside speed control cushions humps," TRL, Crowthrone, Report No. 180 (1995).

[18] P. Kokowski and R. Makarewicz, "Predicted effects of a speed bump on light vehicles noise," Applied Acoustics **67**, 570–579 (2006).

[19] Anon., "Inventory of noise mitigation methods," European Commission, Environment, Working Group 5 (2002).

[20] V. Desarnaulds, G. Monay, and A. Calvaro, "Noise reduction by urban traffic management," *International Congress on Acoustics-2004*, Kyoto.

[21] H. Bendsen, J. Haberl, U. Sanberg, and G. Watts, "Traffic management and noise reducing Pavements," Danish Road Institute, Report No. 137 (2004).

[22] H. Bendsen and L. E. Larsen, "Traffic management and noise," *Inter-Noise 2006*, Honolulu.

[23] S. Kono, Y. Oshino, T. Iwase, T. Sone, and H. Tachibana, "Road traffic noise prediction model ASJ RTN—Model 2003 proposed by the Acoustical Society of Japan, Part 2," *International Congress on Acoustics 2004*, Kyoto.

[24] B. M. Favre, "Noise emission of road vehicles: Evaluation of simple model," J. Sound Vib. **91**, 571–582 (1983).

[25] R. Makarewicz and P. Kokowski, "Simplified model of ground effect," J. Acoust. Soc. Am. **101**, 372–376 (1997).

[26] K. Attenborough, T. Waters-Fuller, K. M. Li, and J. A. Lines, "Acoustical properties of farmland," J. Agric. Eng. Res. **76**, 183–195 (2000).

[27] K. Attenborough, "A comparison of engineering methods for predicting ground effect," *Forum Acusticum*, Berlin (2000).

[28] R. Makarewicz and R. Golebiewski, "Modeling of the roundabout noise impact," J. Acoust. Soc. Am. **122**, 860–868 (2007).

[29] R. Akcelik and M. Besley, "Acceleration and deceleration models," *23rd Conference of Australian Institute of Transport Research*, Melbourne (2001).

Attenuation of high-level impulses by earmuffs

Jan Zera[a)]

*Central Institute for Labour Protection – National Research Institute, Czerniakowska 16, 00-701
Warsaw, Poland, Frederic Chopin Academy of Music, Okólnik 2, 00-368, Warsaw, Poland, and Faculty of
Electronics and Information Technology, Warsaw University of Technology, Nowowiejska 15/19,
00-665 Warsaw, Poland*

Rafal Mlynski[b)]

*Central Institute for Labour Protection – National Research Institute, Czerniakowska 16, 00-701 Warsaw,
Poland*

Attenuation of high-level acoustic impulses (noise reduction) by various types of earmuffs was measured using a laboratory source of type *A* impulses and an artificial test fixture compatible with the ISO 4869-3 standard. The measurements were made for impulses of peak sound-pressure levels (SPLs) from 150 to 170 dB. The rise time and *A* duration of the impulses depended on their SPL and were within a range of $12-400$ μs (rise time) and $0.4-1.1$ ms (*A* duration). The results showed that earmuff peak level attenuation increases by about 10 dB when the impulse's rise time and the *A* duration are reduced. The results also demonstrated that the signals under the earmuff cup have a longer rise and *A* duration than the original impulses recorded outside the earmuff. Results of the measurements were used to check the validity of various hearing damage risk criteria that specify the maximum permissible exposure to impulse noise. The present data lead to the conclusion that procedures in which hearing damage risk is assessed only from signal attenuation, without taking into consideration changes in the signal waveform under the earmuff, tend to underestimate the risk of hearing damage. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2756973]

I. INTRODUCTION

This paper reports a study which was carried out to measure the characteristics of attenuation of impulse noise by various types of earmuffs. The two main variables used to describe the effect of earmuff attenuation on impulses were the peak sound-pressure level (SPL) and the impulse duration. In particular, the peak SPL, the A-weighted energy, the rise time, the *A*, *C*, and *D* durations of impulses were measured under and outside the earmuff.

Although it has been often emphasized that an adequate assessment of the efficiency of a hearing protector[1–3] should be made on the basis of the physiological response of the auditory system, such as the temporary threshold shift (TTS), protection against impulse noise is often determined by objective measurement of insertion or transmission loss. Objective measurements may be made either with the use of artificial test fixtures (ATFs) (Refs. 1 and 4–7) or on human subjects (MIRE technique).[3,6–12] The MIRE technique poses some risk of exposing the subjects to excessive levels of impulse noise even if the subject's ear is additionally protected by an earplug placed under the earmuff.[13] Therefore, the application of MIRE technique is restricted at high levels to measurements made in real situations of exposure, such as military exercises. Only the use of ATF ensures safe mea-surement conditions, and such a method may be used in the laboratory with no limitations as to the impulse levels and the number of exposures.

In published studies conducted with the TTS or the MIRE technique on human subjects in real environments, measurements were made for a very small number of hearing protectors, involving three,[10,14] two,[5,8,9,15,16] or in some cases only one[11,17] hearing protector. Therefore, certain conclusions drawn from these studies, such as the linking of the peak SPL attenuation to specific physical characteristics of the earmuff, should be treated with caution. In contrast, in the present study, measurements of transmission loss were made for the relatively large number of 32 models of earmuffs from various manufacturers.

The measurements were made with the use of the ATF recommended in the ISO 4869-3 standard.[18] Measurements on the ATF are quick enough to collect the data needed for a comparison of the 32 hearing protection devices (HPDs) that were tested. A possible disadvantage of using an ATF is that the assessment of the HPD's attenuation may be less accurate than in measurements made on humans.[1,19] Certain studies have reported that the ATF yields a $10-20$-dB higher estimate of attenuation than the REAT method that employs human subjects.[6] However, in other studies (see Berger[1] for a review) such a large discrepancy was observed only at frequencies above $1-2$ kHz, whereas at lower frequencies the attenuation values determined on the ATF were only higher by 1 or less dB than those obtained using the REAT method.[1,17] Regardless of any limitations as to the relation of ATF measurements to real situations in which HPDs are

---

a)Electronic mail: jazer@ciop.pl
b)Electronic mail: rmlynski@ciop.pl
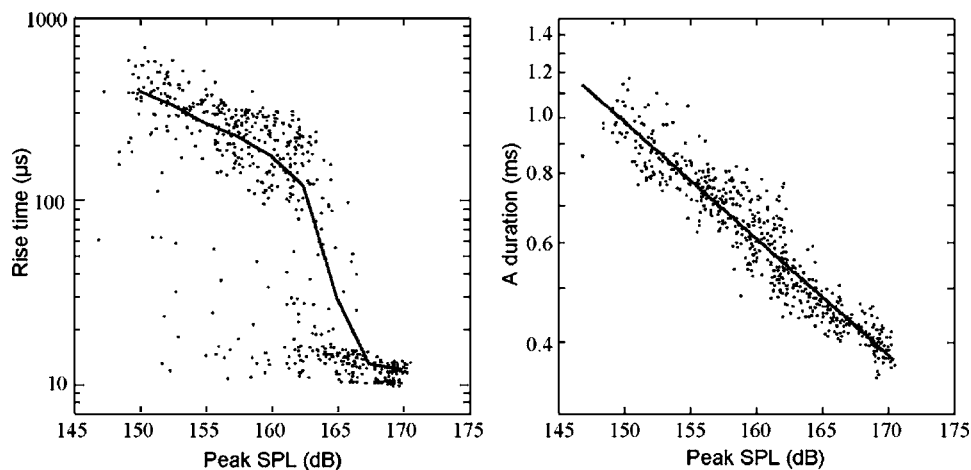
FIG. 1. Rise time (left panel) and *A* duration (right panel) versus peak SPL of impulses outside the earmuff. Solid lines represent the average rise time or the average *A* duration at a given SPL.

used, the data reported in this paper make it possible to assess the performance of various models of HPDs in identical laboratory conditions. In the present study, an attempt was also made to compare the SPLs and impulse durations measured under and outside the earmuff against various hearing damage risk criteria that specify the allowable exposure to impulse noise.[20–26]

II. METHOD

A. Impulse noise source

Acoustic impulses were generated by a custom-made impulse sound source.[27] The main body of the device consisted of a 80-cm-long and 20-cm-wide steel cylinder, rigidly closed at the bottom and covered with an elastic membrane at its top end. The device operated on the principle that release of the air compressed in the cylinder produced an A-type acoustic impulse.[28] of required SPL. The static pressure of air inside the cylinder was precisely controlled to obtain an impulse of the desired peak SPL within a range from about 150 to over 170 dB.

Due to the use of a mechanical technique for generating the impulses, the peak SPL and the *A* duration could not be controlled independently. An increase in peak SPL from 150 to 170 dB was associated with the reduction of the impulse rise time from about 0.4 ms to 12 μs, and the reduction of *A* duration from 1.1 to 0.4 ms (Fig. 1). This effect imposed certain limitations as to the conditions under which the earmuffs could be tested.

B. Measurement setup

The measurements were conducted in a sound-insulated cuboidal room ($9 \times 6 \times 3$ m). The room's walls and ceiling were covered with Rockfon sound-absorbing panels to reduce reflections. Reflections from the floor were reduced by partly covering it with mineral wool panels. The reverberation time of the room ranged from 0.1 s (500–8000 Hz) to 0.25 s at 125 Hz.

Earmuff attenuation was measured by the transmission loss method (noise reduction), with the use of an artificial test fixture (ATF) compatible with the ISO/TR 4869-3:1989 standard.[18] Acoustic isolation of the ATF exceeded 60 dB in 1/3-octave bands, within a frequency range of 63–8000 Hz.

The test impulses were captured by two microphones. One microphone was installed in the ATF to record the signal under the earmuff and the other was placed outside the earmuff, at a distance of 10 cm from the side surface of the ATF, to record the impulse in the open space close to the earcup. The ATF and the external microphone were located 30 cm above the membrane surface (Fig. 2), at a distance of 60 cm from the membrane center. The incidence angle of the shock wave against the external microphone was about 30 deg. The close location of the microphones to the earmuff allowed for the monitoring of the impulses free from wall reflections. In order to reflect a frontal position of the impulse source to a potential hearing protector user, the ATF was oriented such that its main axis was perpendicular to the line connecting the ATF and the impulse noise source.

The impulses were recorded using a $\frac{1}{2}$-in. pressure microphone (Brüel & Kjaer, type 4192) under the earmuff and a high-level $\frac{1}{4}$-in. pressure microphone (Brüel & Kjaer, type 4941) outside the earmuff. The response of the outside microphone was linear up to 184 dB SPL. The signals were conditioned using Brüel & Kjaer type 2669 preamplifiers and a Brüel & Kjaer type 2690 (Nexus) signal conditioning amplifier, and then delivered through a 80-kHz low-pass filter (Tucker Davis Technologies, TDT FT2) to analog/digital

FIG. 2. The measurement setup used for transmission loss (noise reduction) measurements of earmuffs.
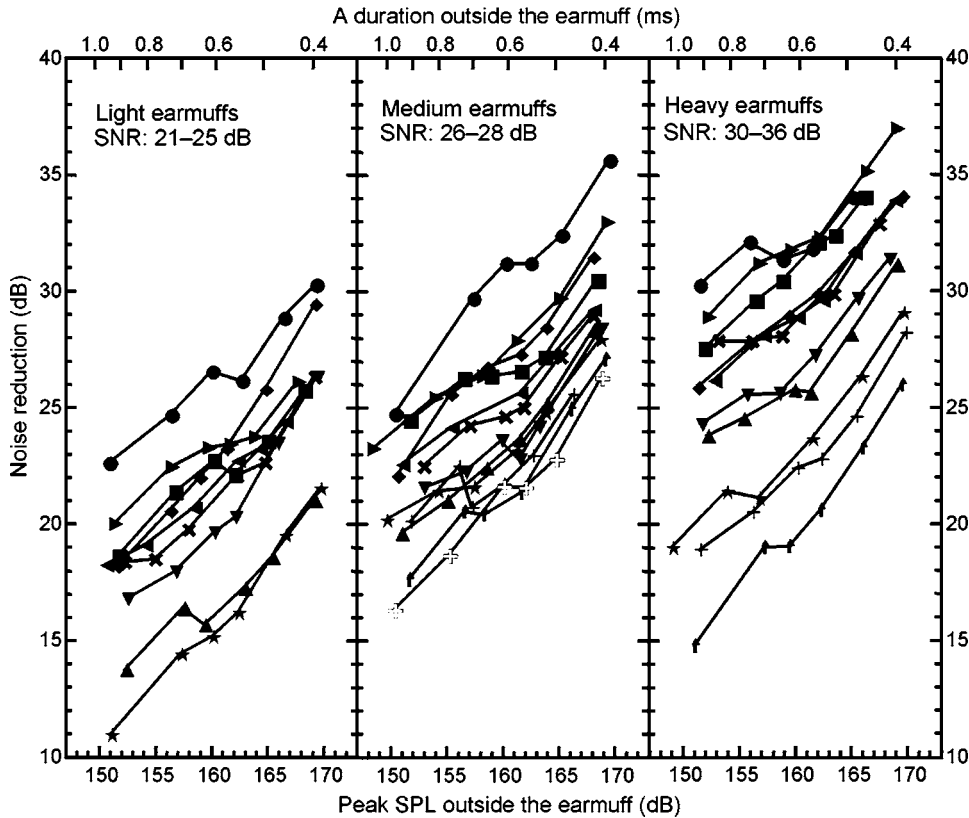
FIG. 3. Reduction of peak SPL of impulses with various *A* duration and peak SPL measured for light (left panel), medium (center panel), and heavy earmuffs (right panel). Secondary axis shows the *A* duration of the impulse outside the earmuff. Each point is the average of three measurements.

(A/D) converters (TDT DD1). A 168-kHz sampling rate was used for A/D conversion, which enabled us to read the time parameters and the peak SPL with sufficient accuracy. Although the sampling rate and the low-pass filter's cutoff frequency permitted A/D conversion up to 80 kHz, due to the frequency response of the microphones, the signal frequency bandwidth was limited to 20 and 40 kHz, respectively, for signals recorded inside and outside the earmuff. The digitized waveforms were stored on a computer disk. Further processing in a MATLAB programming environment was used to calculate the peak SPL, L_{Aeq8}, the rise time, and *A*, *C*, and *D* durations[28] of impulses recorded under and outside the earmuff. Using these values, the decrease in peak SPL, L_{Aeq8} and L_{eq8}, as well as the increase in the impulse rise time and *A*, *C*, and *D* durations, were determined for six values of air pressure in the cylinder of the impulse sound source. By using these settings of air pressure, six levels of impulses were obtained, covering a range of 150–170-dB peak SPL in approximately equal steps. The values of all parameters measured for impulses were determined using the average calculated from the impulse waveforms recorded for three explo-

sions at each air-pressure level in the sound source cylinder. Altogether, over 1150 impulses were measured simultaneously outside and inside the earmuffs.

III. HEARING PROTECTORS

The set of hearing protectors measured in the study was comprised of 32 models made by various manufacturers. All the earmuffs tested are listed in Table I. For maximum clarity in presenting the data, the earmuffs in Table I are ordered according to the single number rating (SNR) values specified by the manufacturers and divided into three categories, labeled "light" (SNR=21–25 dB), "medium" (SNR =26–28 dB), and "heavy" (SNR=30–36 dB). This classification is consistently used in the presentation of data in further sections of the paper.

IV. RESULTS AND DISCUSSION

A. Impulse noise reduction by earmuffs

The measured values of peak SPL attenuation (noise reduction) are shown in Fig. 3, in separate panels for light,

2084   J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

J. Zera and R. Mlynski: Attenuation of high-level impulses by earmuffs

medium, and heavy earmuffs. The main abscissa is the peak SPL of impulses recorded outside the earmuff, and the second abscissa axis shows the corresponding A duration of the impulse. The ordinate is the peak SPL attenuation (noise reduction) by earmuffs. The peak SPL attenuation at 150- and 170-dB peak SPLs of the impulse outside the earmuff, and the average attenuation within the range from 150- to 170-dB peak SPL are also given in Table I.

The data plotted in Fig. 3 show that peak SPL attenuation increases by 6–11 dB when the peak SPL outside the earmuff is increased and the impulse's A duration is reduced. Additional measurements described in Sec. IV B clearly indicate that this increase in attenuation is due to the reduced A duration of the impulse, and is not caused by an increment of SPL outside the earmuff. The decrease in peak SPLs of impulses with shorter A duration and shorter rise time was larger than in the case of impulses with longer A duration and rise time.

Attenuation of peak SPL by light earmuffs (SNR 21–25 dB) amounts to 11–22 dB for 1-ms A duration at 150-dB peak SPL. However, when the impulse's A duration is reduced from 1 to 0.4 ms at 170-dB peak SPL, the effective attenuation of the same earmuffs increases by about 10 dB and ranges from 21 to 30 dB. This effect of signal waveform on peak-level attenuation has been known from previous studies in which impulse noise was produced by explosives and various kind of weapons. Parmentier and Buck[15] used explosives and found that, at a constant peak SPL of 174 dB, the attenuation of peak SPL provided by two combat earmuffs was reduced by about 15 dB when the impulse's A duration increased from 0.15 to 1 ms. Further increases in the A duration to 4.5 ms reduced the attenuation by another 10 dB, which in this case was 25 dB less than for an impulse with 0.15-ms A duration. In contrast, an increase in outside impulse peak SPL from 148 to 174 dB at a constant 2-ms A duration caused only a 9-dB reduction of earmuff attenuation, from 26 to 17 dB, and no reduction of attenuation was observed when peak SPL was further increased to 190 dB. A study by Parmentier and Buck[15] has clearly shown that the influence of impulse A duration on the earmuff attenuation of high-level impulses was stronger than the effect of peak SPL.

For medium earmuffs (Fig. 3, middle panel), the peak SPL of impulses decreases by 16–24 dB, and 26–36 dB, respectively, for an A duration of about 1 ms at 150-dB peak SPL and 0.4 ms at 170 dB. On average, medium earmuffs provide about 5 dB more attenuation than light earmuffs, which is in fairly good agreement with the SNR values specified by the manufacturers.

With a few exceptions, heavy earmuffs (SNR 30–36 dB) provide the most effective attenuation of impulses (Fig. 3, right panel), providing attenuation of 24–29 and 31–37 dB for A durations of 1 and 0.4 ms, respectively. Thus, for an impulse of 1-ms A duration, heavy earmuffs provide 5–8-dB higher attenuation than light and medium earmuffs. An exception in this category of earmuffs, the reason for which is yet unknown to us, is reflected in the results for the JSP Big Blue, the JSP Goodwood, and the Bilsom Marksman Pro earmuffs, which attenuate the peak SPL by

only 15–19 dB for A duration of 1 ms, or 26–29 dB for A duration of 0.4 ms, in spite of their high SNR of 29–31 dB.

An insertion loss of the ATF used for the measurements amounts to 60 dB. In real conditions of earmuff use the bone conduction paths may limit the attenuation, at least for those few models which display attenuation of 35 dB and more for impulses at 170-dB peak SPL and short A duration. Peak SPL attenuation of 35 dB is about 5 dB lower than the minimum value of the bone conduction limit of 40–42 dB at 2 kHz. Since the impulse spectrum level at 2 kHz is about 10–15 dB below the peak SPL, it may be expected that the bone conduction does not reduce the earmuff's attenuation significantly.

The increase in the earmuff's SNR may be linked to the increase in the cup volume and mass. For instance, Pekkarinen et al.[10] measured attenuation of three models of earmuffs that were also tested in the present study—Silenta Mil, Silenta Ergo, and Silenta Super—in conditions of exposure to impulse noise produced by antitank weapons. These models, classified in their study as small-, medium-, and large-size earmuffs, differed in the cup volume. Pekkarinen et al.[10] observed that earmuff attenuation of peak SPL correlated with the cup volume. The 4–12-dB increase in attenuation reported by Pekkarinen et al.[10] is in close agreement with the 5–8-dB increase in peak-level attenuation found in the present study.

The data plotted in Fig. 3 also show that, depending on the A duration, the attenuation of peak SPL may be either lower or higher than the SNR value specified by the manufacturer. For 1-ms A duration, the decrease in peak SPL in most earmuffs is 2–10 dB less than the SNR value, and in the case of a few models of heavy earmuffs this difference reaches 15 dB. However, for 0.4-ms A duration, the average decrease in peak SPL exceeds the SNR value by only 1.8 dB. For practical reasons, it may be useful to determine the impulse A duration for which the SNR value most accurately predicts the difference in the peak SPL measured outside and under the earmuff. For most earmuffs, the A duration for which the SNR value most accurately reflects the reduction in peak SPL is within the range of 0.4–0.5 ms, whereas for impulses with longer A duration the SNR parameter overestimates the reduction in peak SPL. The decrease in attenuation for long A durations may be attributed to the less effective attenuation of impulse energy at low frequencies caused by the low-pass frequency characteristics of earmuff attenuation.

Similar analysis was conducted for the H, the M, and the L parameters[29] which represent the hearing protectors' of high-, mid- and low-frequency attenuation. The H parameter most accurately reflects the reduction in peak SPL of impulses with A duration of about 0.4 ms, the M parameter the reduction in peak SPL at A duration of 0.6 ms, and the L parameter the reduction in peak SPL at A duration of 0.9 ms. This result is in good agreement with an approximate method for assessing the hearing protector attenuation of impulsive noise described in Annex B in the EN 458 standard[30] in which the peak SPL attenuation is predicted based on the H, M, and L parameters for short, medium, and long A-duration impulses, respectively.
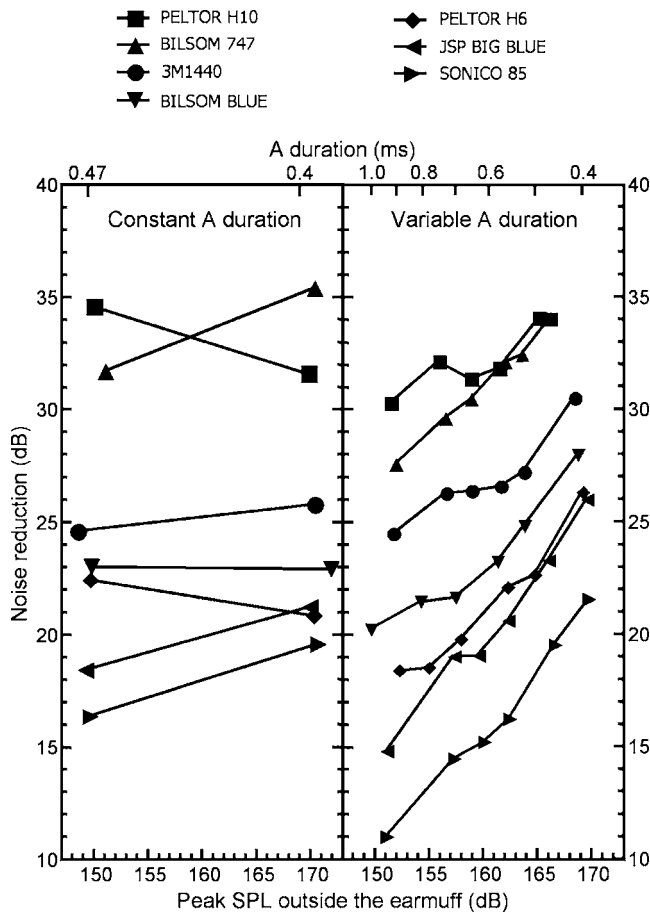
FIG. 4. Reduction of peak SPL of impulses at different peak SPL and constant (left panel) and variable (right panel) A duration.

The published data on the attenuation of impulse noise by the models of hearing protectors listed in Table I are sparse. For the Peltor H7 earmuff, the average noise reduction measured in the present study was 29.7 dB. For this model, Bronkhorst and Smoorenburg[5] reported 27.3-dB attenuation for impulses generated by a drum (a value corrected for an open-ear condition) and about 37-dB attenuation for pistol shots. The high attenuation found for pistol shots was due to the presence of intense high-frequency components in the frequency spectrum. The results obtained by Smoorenburg[6] for the Peltor H7 earmuff, using the MIRE technique, have demonstrated a 27–40-dB attenuation of peak SPL for various industrial impulses. As the industrial impulses in Smoorenburg's study[6] considerably varied in duration and frequency spectrum (e.g., a punch press: low-frequency content and short rise time; a nail pistol: high-frequency content and short rise time), the differences in attenuation observed for these types of noise were substantial. When explosives were used, the attenuation provided by the Peltor H7 earmuff was only 17 dB due to large pulse durations (12–15 ms).[6]

A decrease in attenuation caused by an increase in impulse A duration was also reported by authors who studied protection from impulses produced by large-caliber weapons with long A durations. Ylikoski et al.[11] measured the attenuation provided by the Silenta Mil earmuff for pistol and rifle shots of 0.3-ms A duration and for a 130-mm cannon shot of 7-ms A duration. The decrease in peak SPL was 30 dB for pistol shots and 16 dB for rifle shots, whereas in the case of a cannon shot the peak SPL decreased by only about 4 dB. The finding that the low attenuation of impulses with long A durations is an effect of poor attenuation at low frequencies is supported by the results of various published experiments. The present data are also in agreement with results of Ylikoski et al.[11] As seen in Table I, noise reduction for the Silenta Sport Mil Plus earmuff was 28.4 dB for an impulse with 0.4-ms A duration and 21.6 dB for an impulse with 1-ms A duration. In another study Ylikoski et al.[8] measured the attenuation of impulses encountered in the military by both small (Silenta Mil) and large (Silenta Super) cup volume earmuffs. Ylikoski et al. reported that the attenuation of peak SPL by an earmuff with a small cup volume was 29 dB for a pistol shot (0.3-ms A duration), 17.4 dB for a bazooka shot (3.2-ms A duration), and 7.7 dB for a 130-mm-cannon shot (4.6-ms A duration). They also found that an earmuff with a large-volume cup provided 10 dB more attenuation for impulses than a small-volume earmuff. In contrast, the difference in the protection afforded by light and heavy Silenta earmuffs as determined in our study did not exceed 5 dB (see Fig. 3 and Table I). Attenuation of heavy weapon shots by an earmuff with a large-volume cup measured by Ylikoski et al.[8] was 10–15 dB less than the attenuation values obtained for several models of Silenta earmuffs in our study (Table I). This difference was possibly due to the fact that the A duration of weapon shots used by Ylikoski for the measurement of attenuation was about 4 times longer than the A duration of 0.4–1 ms of the impulse in our study.

It is important to note that the reduction of peak SPL determined with the use of an ATF in the present study is larger than that determined in measurements performed on human subjects with the use of the MIRE technique. Starck et al.,[14] who employed the MIRE technique, reported that the peak SPL attenuation of industrial impulse noise was 8–14 dB for small-volume cup earmuffs, whereas for medium and large-volume cup earmuffs it was 12–25 dB. They also found that the reduction of impulse energy by earmuffs with small, medium, and large cup sizes was 6–12, 13–20, and 19–24 dB, respectively. These values of attenuation, especially for light- and medium-size earmuffs, are about 5–10 dB lower than those obtained in the present study (Figs. 3 and 11).

As mentioned earlier, there is some deviation between data based on the MIRE method and those obtained with the use of an ATF.[19] For instance, Pekkarinen et al.,[10] who used the MIRE technique, obtained for a 171-dB peak SPL impulse produced by a hand-held bazooka only a 9–15-dB decrease in peak SPL under the Silenta Mil, Silenta Ergo, and Silenta Super earmuffs, which differed in cup mass and volume. Attenuation measured for these earmuffs in our study was 28–35 dB. The higher values of transmission loss obtained in our study are partly due to the measurement method[1] and to the difference in the frequency content of the impulse generated by the laboratory source. There are two reasons why the use of ATF yields higher attenuation values than the MIRE technique. First, the earmuff seal is tighter on the flat surface of the ATF than on the curved surface of the

FIG. 5. Rise time of impulses measured outside and under the earmuff, for light (left panel), medium (center panel), and heavy earmuffs (right panel). Secondary axis shows the peak SPL of the impulse outside the earmuff. Each point is the average of three measurements.

side of a human skull. Second, the soft tissue of a real head and ear canal increases the compliance of the system comprised of the cup, cushion, and air volume.

B. Noise reduction at constant A duration

It has been commonly accepted that earmuff response to a shock wave is linear within a range of peak SPL not exceeding 150 dB, and that nonlinear effects are not clearly evident for peak SPL below 170 dB.[26,31] When the earmuff response is considered linear, the effect of a decrease in earmuff attenuation with increasing A duration is attributed to the increased low-frequency energy in the impulse spectrum.[6,8,10,11] For this reason, the change of earmuff attenuation with peak SPL seen in Fig. 3 may be linked to the decrease in the A duration (as a linear effect) rather than to the actual increase in peak SPL (a possible nonlinear effect).

An additional experiment was conducted to verify whether the peak SPL or the A duration of the impulse outside the earmuff could cause a change in earmuff attenuation. In this experiment, the peak SPL of the impulse reaching the earmuff was controlled independently of the impulse's A duration by increasing the distance between the impulse noise source and the ATF.

The measurements of peak SPL under and outside the earmuff were made at distances of 0.5 and 6 m between the impulse source and the ATF. At a distance of 0.5 m, the peak SPL of impulses reaching the earmuff was about 170 dB. When the distance was increased to 6 m the peak SPL decreased to about 150 dB, with very little change in the impulse's A duration, which was 0.4 ms at 170 dB and 0.47 ms at 150 dB. The volume of the measurement chamber was too small to test the earmuffs in these conditions; therefore, the measurements were made on a grassy field outside the laboratory.

The decrease in peak SPL determined for seven earmuffs, at distances of 0.5 or 6 m between the impulse source and the ATF, is shown in the left panel of Fig. 4. For comparison, the right panel of Fig. 4 presents the results obtained for the same earmuffs in conditions where the peak level was controlled by changing the air pressure inside the cylinder, which caused a change in both peak SPL and A duration of
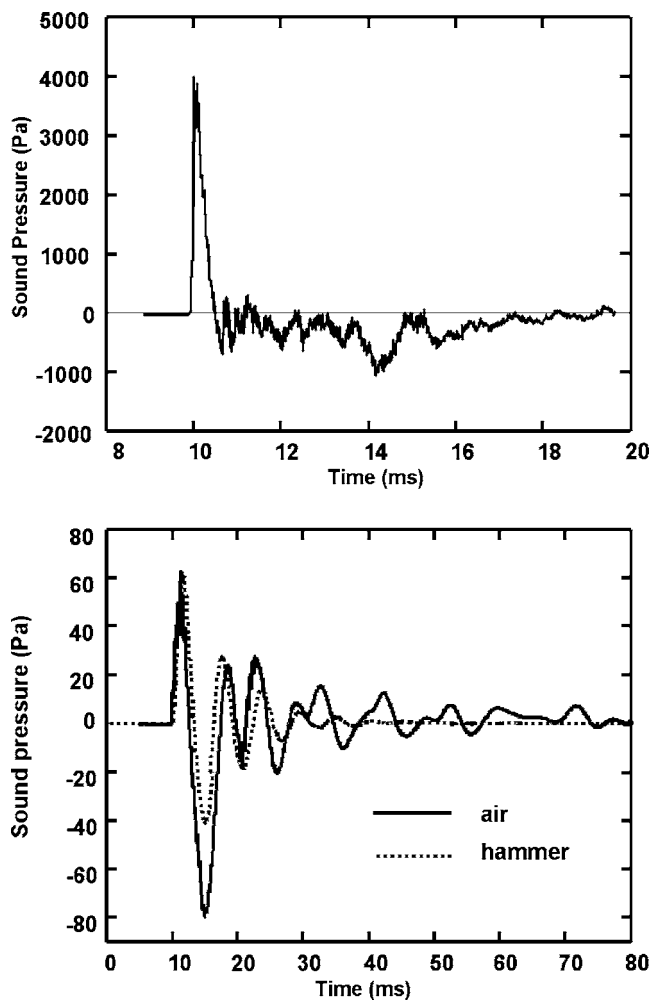
FIG. 6. Upper panel: the waveform of the impulse recorded outside the earmuff. Lower panel: the waveform of the impulse recorded under the earmuff, produced by the impulse source or by striking the cup with a rubber hammer.

the impulse outside the earmuff (data replotted from Fig. 3). In conditions with a constant A duration of the outside impulse, at peak SPLs 150 or 170 dB, the variation of attenuation of peak SPL by earmuffs was reduced to about ±3 dB. This result is in contrast with measurements reported in Fig. 3, in which a constant increase by about 10 dB in earmuff's attenuation was observed when both the A duration and the peak level of the impulse were increased. The results plotted in Fig. 4 clearly show that the increase in earmuff attenuation with peak SPL from 150 to 170 dB should be attributed to the decrease in A duration of the impulses.

C. Effect of earmuff attenuation on the rise time and A, C, and D duration of the impulse

There is a significant difference between an A-type waveform of an impulse in open space and the waveform recorded under the earmuff. One of the effects is that the rise time of the impulse is longer under the earmuff than outside the earmuff.[31,32] Moreover, a noteworthy effect is that the values of rise time measured under and outside the earmuff are not linearly related, as is apparent in Fig. 5.

The ordinate in Fig. 5 shows the rise time of impulses recorded under the earmuffs for the hearing protectors listed in Table I, and the abscissa is the rise time of impulses recorded outside the earmuff. The rise time was defined as the time interval within which the instant pressure increased from 10% to 90% of peak value. When peak SPL was reduced from 170 to 150 dB, the rise time of impulses outside the earmuff increased from about 12 μs to about 0.4 ms (Fig. 1, left panel). For each earmuff, the data shown in Fig. 5 were approximated by a best-fitting linear function for values of rise time obtained at six peak SPLs within a range of 150–170 dB.

At peak SPL of 170 dB, the 12-μs rise time of the outside impulse increased under various types of earmuffs by one or two orders of magnitude and ranged from 0.3 to 1.5 ms. At 150-dB peak SPL, the 0.5-ms rise time increased to 0.7–1.5 ms, a range similar to that observed for a 170-dB impulse outside the earmuff. The substantial change, by a factor of about 40, in the impulse's rise time outside the earmuff results in a much smaller, 20%–50%, variation in rise time under the earmuff. The only exception to such a relationship was observed for the Bilsom 747 earmuff (Fig. 5, steep line), but at the present time we are unable to provide an explanation for this discrepancy.

The much longer rise time observed under the earmuff is most likely an effect of the low-pass frequency characteristics of earmuff attenuation.[31,32] The earmuff's frequency response and the time parameters are influenced by various factors, such as the earmuff movement resulting from the shell's mass, the tension of the headband, the compliance of the seal, and the cup volume. Our observations based on striking the earcup with a rubber hammer have shown that the pressure changes under the earmuff are to a large extent similar to the variations caused by an acoustic impulse. The solid line and the dashed lines in Fig. 6 (lower panel) represent the pressure variations under the earmuff after an excitation by the acoustic impulse shown in the upper panel in Fig. 6, and after a hammer strike. It is apparent that during the initial 30 ms of the impulse both pressure waveforms are similar, proving that the earmuff, considered as a spring-mass system, determines the pressure variation under the earmuff.

The data shown in the three panels in Fig. 5 demonstrate that the rise time of impulses recorded under the earmuff generally increases with the SNR value. For light earmuffs (Fig. 5, left panel), for which the SNR value is less than 25 dB, the impulse's rise time ranges from about 0.3 to 1.2 ms under the earmuff. For medium (middle panel) and heavy (right panel) earmuffs, the rise time is within a range of 0.6–1.5 ms.

The effect of earmuff attenuation on the rise time of the impulse reported by other authors is quantitatively in close agreement with the present data. Pekkarinen et al.,[10] who used an impulse with a rise time of 0.2–0.3 ms produced by light hand-held and heavy antitank weapons, reported that the impulse's rise time increased to 1–2.5 ms under the earmuff. The present data show a similar rise time under the earmuff, regardless of the rise time measured outside the earmuff, and it should be noted that the ratio of rise time

FIG. 7. The values of A duration of impulses under and outside the earmuff, for light (left panel), medium (center panel), and heavy earmuffs (right panel). Secondary axis shows the peak SPL of the impulse outside the earmuff. Each point is the average of three measurements.

inside and outside the earmuff is not a constant parameter that could be used for describing an earmuff's characteristics.

Figure 7 presents the A duration of impulses recorded outside and under the earmuff. The data show that the A duration under the earmuff ranges from about 1.9 to 4.5 ms, whereas the range of A duration for the outside impulse is 0.4–1.1 ms (see also Fig. 1). The values of rise time and A duration plotted in Figs. 5 and 7 are also given in Table II. It should be noted that the waveform under the earmuff differed from the Friedlander wave[33] for which the A duration is defined. For this reason, the A duration of the impulse under the earmuff was calculated as the width of the waveform's first peak.

Figures 8 and 9 show the C and D durations, respectively, of impulses recorded under the earmuff. Similarly as in Figs. 5 and 7, these data are plotted against the A duration of the outside impulse. The C duration has been defined, according to Pfander,[21] as the total duration of intervals during which the signal waveform exceeded the −10-dB level, relative to the peak value. As seen in Fig. 8, the values of C

duration range from 5 to 40 ms for most of the earmuffs tested. A comparison of data in the left, middle, and right panels in Fig. 8 shows that the impulse's C duration is independent of the size of the earmuff expressed by the SNR value. In general, the values of C duration under the earmuff are higher for long A durations and low peak SPLs of an outside impulse. Similarly to what was observed for rise time, the A and C durations measured in the present study are in close agreement with the data obtained by Pekkarinen et al.,[10] who reported A durations of about 5 ms and C durations of 5–17 ms for impulses produced by antitank weapons. In contrast to their data,[10] we found no straightforward relation between the A, C, or D durations and the earmuff size.

The values of D duration, calculated for waveforms under the earmuff (Fig. 9), were defined, according to Smoorenburg's criterion,[34] as the time interval needed for the signal envelope to drop below the −10-dB level, relative to the peak value. As might be expected, the values of D duration are higher than those of C duration and range from 5 to

TABLE I. Earmuffs tested in the study. Data include the SNR values specified by manufacturers and decrease in peak SPL, L_{Aeq8} and L_{eq8} under conditions labeled as I, II, and III.

	SNR (dB)	L_{peak} (dB)			L_{Aeq8} (dB)			L_{eq8} (dB)		
		I[a]	II[b]	III[c]	I	II	III	I	II	III
Light earmuffs, SNR: 21–25 dB										
SONICO 85	21	11.0	16.3	21.6	12.7	17.9	21.4	1.5	4.0	5.9
SILENTA BELL II	21	18.2	21.4	24.4	17.6	21.3	23.7	10.9	12.1	12.6
SONICO STANDARD	23	13.8	17.1	21.0	14.5	16.6	17.7	5.6	6.4	6.9
JSP JMUFF	24	18.6	22.4	25.7	19.7	22.6	24.5	10.3	11.4	12.0
OPTA OS-5N	24	16.8	20.8	26.3	20.4	23.4	26.3	11.0	10.1	11.3
PELTOR H6A	24	18.4	21.3	26.3	19.0	21.0	23.4	10.5	11.0	11.8
SILENTA SUPERMIL	25	20.0	23.2	26.1	22.4	24.4	25.5	12.6	13.0	13.1
3M/1435	25	22.6	26.5	30.3	26.9	29.6	30.9	16.3	16.6	16.5
BILSOM POCKET	25	18.2	23.2	29.4	19.5	23.4	26.3	9.6	11.5	13.1
Medium earmuffs, SNR: 26–28 dB										
BILSOM LOTON 2401	26	22.6	25.8	29.2	22.8	26.5	27.4	16.0	17.3	16.3
PELTOR H9A	26	16.3	21.2	26.3	18.9	22.8	25.4	9.2	11.1	12.5
SILENTA UNIVERSALL	26	23.3	27.6	33.0	25.2	27.1	29.2	16.2	16.3	16.8
EAR 1000	27	17.6	22.0	27.1	20.6	23.8	25.5	10.4	11.8	12.7
SILENTA SPORT MIL PLUS	27	21.6	23.8	28.4	23.7	25.7	27.0	12.9	13.6	14.0
3M/1440	27	24.5	26.9	30.5	26.2	29.0	31.1	15.1	15.5	16.0
BILSOM 717	27	20.1	22.3	25.6	21.5	23.5	25.0	12.5	13.0	13.3
SILENTA ERGO II	27	24.7	30.8	35.6	26.1	31.0	33.2	15.8	17.8	19.1
A 812 X	27	19.6	23.4	28.4	22.2	25.3	27.3	12.6	14.2	15.4
PELTOR H3A	28	22.5	25.4	29.0	25.4	27.4	28.7	19.0	16.6	15.7
SILENTA SUPER	28	22.1	26.9	31.5	23.5	28.9	32.5	13.9	16.4	18.2
BILSOM BLUE 2450	28	20.2	23.2	28.0	21.5	22.8	24.8	13.8	13.8	14.4
Heavy earmuffs, SNR: 30–36 dB										
BILSOM 727	30	23.8	26.5	31.2	23.6	26.7	28.7	14.5	15.4	16.0
BILSOM 737	30	24.3	27.3	31.4	25.4	27.5	28.9	16.6	17.0	17.4
JSP BIG BLUE	30	14.8	20.4	26.0	15.0	21.6	26.3	4.2	7.0	9.5
BILSOM 747	30	27.5	31.0	34.0	31.3	34.1	35.9	18.4	19.1	19.5
PELTOR H7A	30	25.9	29.7	34.1	28.4	30.7	32.1	18.0	18.7	18.9
BILSOM VIKING 2421	31	26.2	29.7	33.9	28.3	30.6	32.2	19.1	19.8	21.0
BILSOM MARKSMAN PRO 2902	31	19.0	23.4	29.1	23.6	26.2	28.4	13.8	14.5	15.2
JSP GOODWOOD	31	18.9	22.9	28.2	21.1	25.2	28.7	8.9	10.1	11.6
JSP MONZA	31	27.9	29.4	32.9	30.0	32.1	33.6	19.6	19.3	19.8
PELTOR H10A	34	30.3	32.3	34.0	31.7	33.7	35.0	23.1	23.1	23.2
JSP MONACO	36	28.9	32.7	37.0	30.0	31.7	32.0	21.2	22.3	23.1

[a]I - at 150 dB (mean A duration=0.90 ms).
[b]II - average in 150–170-dB range.
[c]III - at 170 dB (mean A duration=0.42 ms).

75 ms, with the majority of values below 50 ms. As seen in Fig. 9, D duration values are related neither to the type nor the size of the earmuff.

V. EARMUFF ATTENUATION IN RELATION TO DAMAGE RISK CRITERIA

Hearing damage risk criteria concerning exposure to impulse noise (CHABA,[23] Pfander,[21,22] Smoorenburg,[34] and Dancer[20,25]) have been developed for impulses encountered in free-field conditions; therefore, the applicability of these criteria to signals measured under HPDs has been questioned.[25] The data obtained in this study make it possible to assess the relevance of hearing damage risk criteria to the conditions in which hearing protectors are used. This can be done by applying the above criteria to the parameters of the impulses recorded under the hearing protector.

Figure 10 shows the CHABA, the Pfander, and the Smoorenburg hearing damage risk criteria applied to the present data. The filled and open circles in all panels in Fig. 10 refer to impulses recorded outside and inside the earmuff, respectively. The left panel shows the peak SPL of impulses plotted against A duration; the line indicates the CHABA criterion[23] recommended for conditions of exposure to a single impulse. The application of CHABA criterion implies that the majority of impulses recorded outside the earmuff (filled circles), and all impulses recorded under the earmuff (open circles) may be regarded as safe for hearing.

Pfander[21,22] proposed a simple linear correction to adjust the hearing damage risk criteria to conditions in which HPDs

TABLE II. Rise time and *A* duration of impulses measured under the earmuff at 150- and 170-dB peak SPL of the impulse outside the earmuff.

	Rise time under the earmuff at 150-dB peak SPL and 345-μs mean rise time outside the earmuff (ms)	Rise time under the earmuff at 170-dB peak SPL and 14-μs mean rise time outside the earmuff (ms)	A duration under the earmuff at 150-dB peak SPL and 0.9-ms mean A duration outside the earmuff (ms)	A duration under the earmuff at 170-dB peak SPL and 0.4-ms mean A duration outside the earmuff (ms)
	Light earmuffs, SNR: 21–25 dB			
SONICO 85	0.74	0.32	3.9	2.9
SILENTA BELL II	1.12	1.06	2.7	2.3
SONICO STANDARD	0.68	0.48	2.9	2.6
JSP JMUFF	0.97	0.94	3.0	2.7
OPTA OS-5N	1.13	1.05	3.7	3.1
PELTOR H6A	0.98	0.79	2.8	2.4
SILENTA SUPERMIL	1.12	1.13	3.0	2.7
3M/1435	0.98	0.97	3.3	2.9
BILSOM POCKET	0.74	0.48	3.6	1.9
	Medium earmuffs, SNR: 26–28 dB			
BILSOM LOTON 2401	0.92	0.86	2.6	2.3
PELTOR H9A	1.26	0.97	3.4	2.6
SILENTA UNIVERSALL	0.92	0.65	2.8	2.6
EAR 1000	1.03	0.90	3.2	2.7
SILENTA SPORT MIL PLUS	1.18	0.96	3.2	2.7
3M/1440	1.02	0.83	3.4	2.8
BILSOM 717	0.98	0.91	2.9	2.5
SILENTA ERGO II	1.30	1.19	3.2	2.8
A 812 X	1.16	0.91	3.3	2.6
PELTOR H3A	1.11	1.06	3.2	2.9
SILENTA SUPER	1.21	0.68	3.6	2.8
BILSOM BLUE 2450	0.84	0.69	2.8	2.3
	Heavy earmuffs, SNR: 30–36 dB			
BILSOM 727	1.10	1.16	3.1	2.7
BILSOM 737	0.97	0.98	2.8	2.5
JSP BIG BLUE	1.51	1.28	4.1	3.6
BILSOM 747	2.09	0.09	4.8	3.9
PELTOR H7A	1.20	1.20	3.4	3.0
BILSOM VIKING 2421	0.97	0.99	3.0	2.7
BILSOM MARKSMAN PRO 2902	1.09	1.09	3.3	2.7
JSP GOODWOOD	1.12	1.26	4.2	3.5
JSP MONZA	1.17	1.34	3.7	3.2
PELTOR H10A	1.03	1.14	2.9	2.5
JSP MONACO	1.02	0.79	2.9	2.2

are used. According to this correction, if an HPD is worn the peak SPL recorded in open space should be lowered by a value such as the SNR, representing the HPD's attenuation, typically by 25 dB. It should be noted, however, that this correction gives only a general approximation of peak SPLs under the HPD and does not include the substantial alteration of the impulse's waveform under the earmuff. Our data, with such a correction applied, are shown by crosses in all panels in Fig. 10. For *A* duration (Fig. 10, left panel), corrected values as well as the values measured under the earmuff (open circles), range from 15 to more than 40 dB below the CHABA criterion line, which would mean that the impulses do not cause any risk of hearing damage. However, the data determined with the use of such correction underestimate the

A-duration values measured under the earmuff, and in certain cases also the peak SPLs. This has little influence on the risk assessed using the CHABA criterion (Fig. 10, left panel), but substantial influence when Pfander's[21,22] and Smoorenburg's[34] criteria are used (the middle and the right panels).

The solid line in the middle panel (Fig. 10) represents the Pfander criterion of hearing damage risk,[21,22] which is the maximum permissible peak SPL as a function of *C* duration. The peak SPLs of all impulses recorded outside the earmuff (filled circles) exceed by 15–25 dB the levels defined by the criterion, which indicates that the Pfander criterion is much more stringent than the CHABA criterion. The prediction made by adjustment of SPL (Ref. 21) suggests that most of
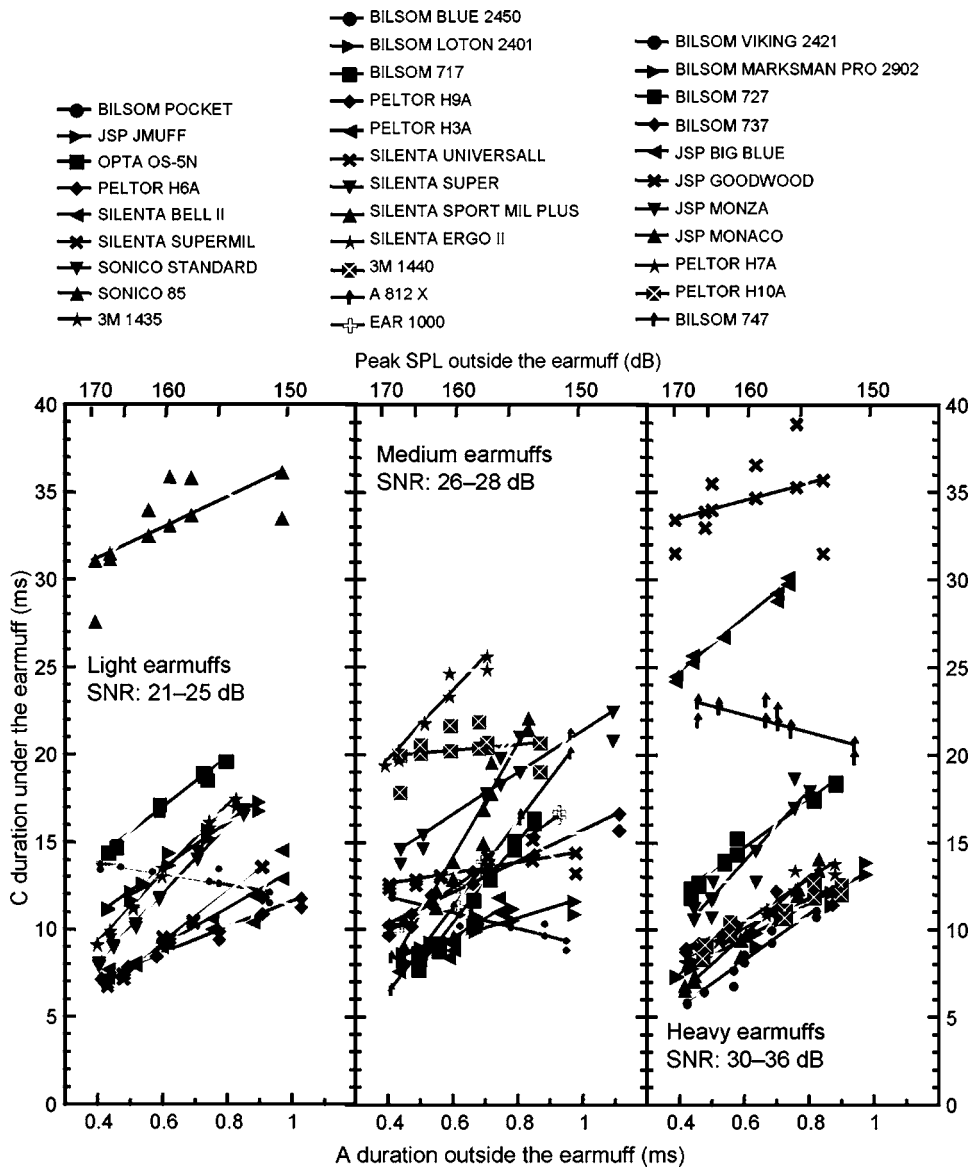
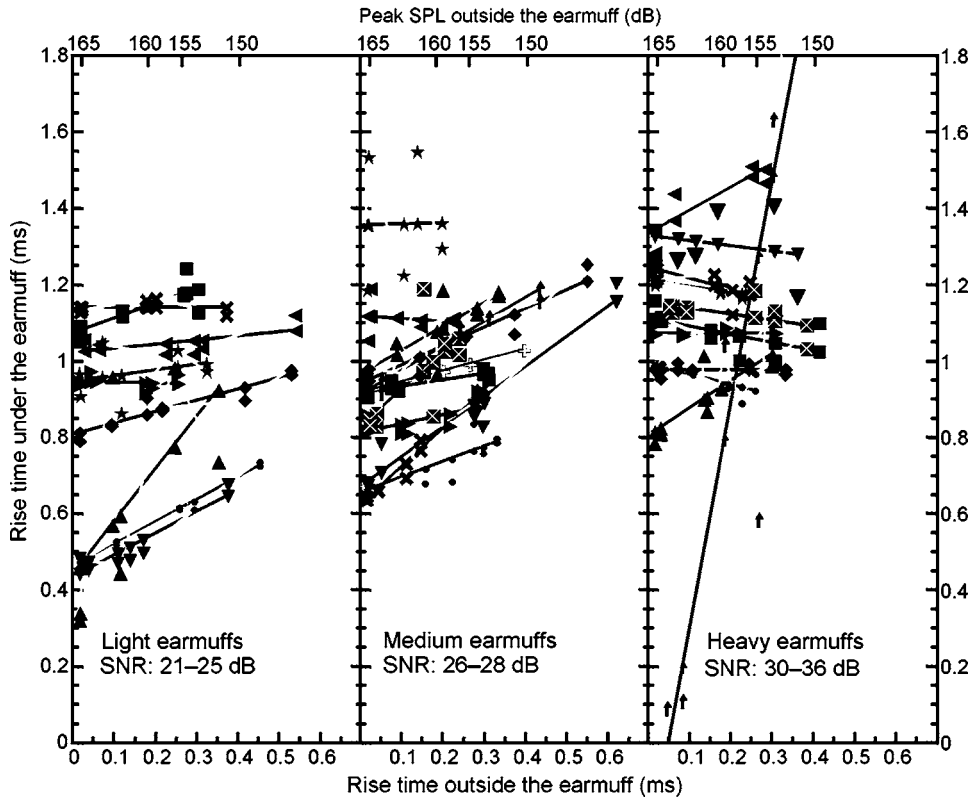FIG. 8. The values of C duration of impulses under the earmuff as a function of the A duration of the impulse outside the earmuff, for light (left panel), medium (center panel), and heavy earmuffs (right panel). Secondary axis shows the peak SPL of the impulse outside the earmuff. Each point is the average of three measurements.

the impulses represented in Fig. 11, middle panel (crosses) would not pose any risk of hearing damage to a hearing protector user. This prediction does not agree with the measurements of SPL under the earmuff (open circles). Although the peak SPLs measured under the earmuff and those estimated from the SNR values fall into the same range, most of the data points obtained from measurements under the earmuff lie above the line representing the Pfander criterion. This results from the fact that under the earmuff the impulses' C duration was about one order of magnitude longer than outside the earmuff, and the maximum permissible SPL was lower than for an impulse with longer C duration.

The solid line in the right panel (Fig. 10) represents the hearing damage risk criterion proposed by Smoorenburg.[34] The data representing the peak SPL and D duration of the impulses under the earmuff lie well above the line indicating the hearing damage risk criterion, an effect also seen in the middle panel, with regard to Pfander's criterion. Applying Pfander's and Smoorenburg's criteria to our data shows that

an arbitrary correction of SPLs, without considering the time characteristics of the waveform under the earmuff, underestimates the possible hearing damage risk posed by exposure to impulse noise.

Dancer[20,25] proposed a damage risk criterion based on the A-weighted 8-h equivalent daily energy exposure level, L_{Aeq8}, to assess the exposure to impulse noise. This criterion makes use of an equal-energy principle, which is a standard for assessing the damage risk from exposure to continuous noise. Therefore, an advantage of using the L_{Aeq8} criterion is that the damage risk resulting from the exposure to impulse noise and that resulting from exposure to continuous noise may be assessed in the same measurement. This approach may be especially useful both in the military and in those industries where impulse noise and high-level continuous noise often occur simultaneously. The energy-based criterion was criticized[11] because it does not take into account the fine temporal structure of the impulse. Nevertheless, Dancer et al.[20,25] argued that the L_{Aeq8} parameter is a proper estimate of

Legend:
- BILSOM BLUE 2450
- BILSOM LOTON 2401
- BILSOM 717
- PELTOR H9A
- PELTOR H3A
- SILENTA UNIVERSALL
- SILENTA SUPER
- SILENTA SPORT MIL PLUS
- SILENTA ERGO II
- 3M 1440
- A 812 X
- EAR 1000
- BILSOM VIKING 2421
- BILSOM MARKSMAN PRO 2902
- BILSOM 727
- BILSOM 737
- JSP BIG BLUE
- JSP GOODWOOD
- JSP MONZA
- JSP MONACO
- PELTOR H7A
- PELTOR H10A
- BILSOM 747
- BILSOM POCKET
- JSP JMUFF
- OPTA OS-5N
- PELTOR H6A
- SILENTA BELL II
- SILENTA SUPERMIL
- SONICO STANDARD
- SONICO 85
- 3M 1435
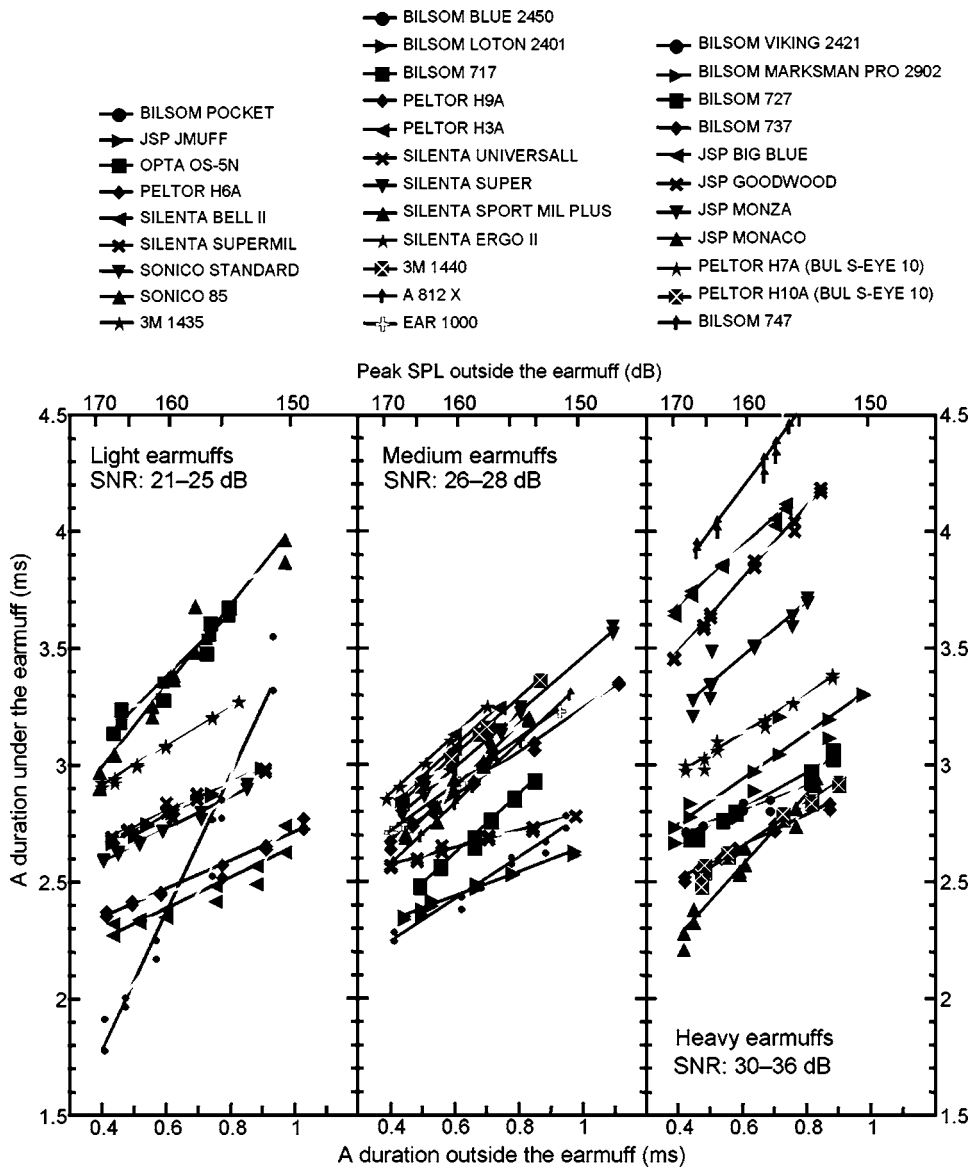
FIG. 9. The values of D duration of impulses under the earmuff as a function of the A duration of the impulse outside the earmuff, for light (left panel), medium (center panel), and heavy earmuffs (right panel). Secondary axis shows the peak SPL of the impulse outside the earmuff. Each point is the average of three measurements.

hearing protector efficiency with regard to impulse noise, and is less stringent than the CHABA, Pfander, and Smoorenburg criteria discussed above.

Figure 11 shows the decrease in L_{Aeq8} level determined in transmission loss measurements. The L_{Aeq8} of impulses outside the earmuff is shown on the abscissa, and the de-
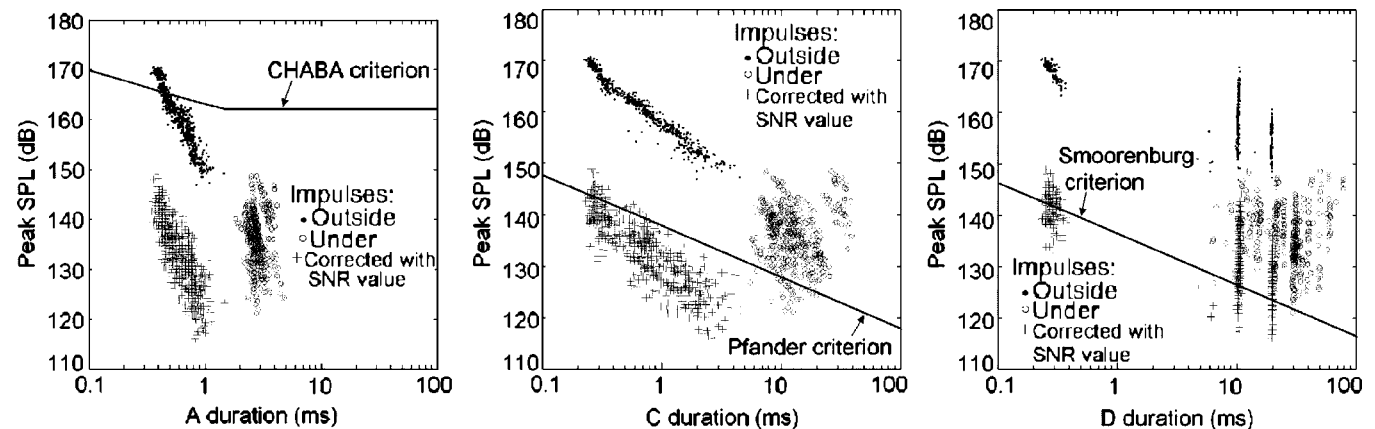
FIG. 10. Peak SPL versus A, C, and D duration (left, middle, and right panel, respectively) for impulses recorded outside (filled circles) and under (open circles) the earmuff. Crosses represent peak SPL measured outside the earmuff and corrected with the SNR value. Solid line shows the maximum permissible level according to the CHABA criterion (left panel), Pfander criterion (middle panel), and Smoorenburg criterion (right panel). The measurements were made for 32 earmuffs.

Legend:
- 3M 1435
- SILENTA SUPERMIL
- JSP JMUFF
- BILSOM POCKET
- SILENTA BELL II
- PELTOR H6A
- OPTA OS-5N
- SONICO STANDARD
- SONICO 85
- SILENTA ERGO II
- SILENTA UNIVERSALL
- 3M 1440
- SILENTA SUPER
- BILSOM LOTON 2401
- PELTOR H3A
- SILENTA SPORT MIL PLUS
- A 812 X
- BILSOM BLUE 2450
- BILSOM 717
- EAR 1000
- PELTOR H9A
- PELTOR H10A
- JSP MONACO
- BILSOM 747
- PELTOR H7A
- BILSOM VIKING 2421
- JSP MONZA
- BILSOM 737
- BILSOM 727
- BILSOM MARKSMAN PRO 2902
- JSP GOODWOOD
- JSP BIG BLUE

FIG. 11. Reduction of L_{Aeq8} for impulses of various A duration and peak SPL, for light (left panel), medium (center panel), and heavy earmuffs (right panel). Secondary axis shows the peak SPL of the impulse outside the earmuff. Each point is the average of three measurements.

crease in the L_{Aeq8} for signals under the earmuff is the ordinate. In order to facilitate comparison with the data shown in Fig. 3, the secondary axis shows the approximate values of peak SPLs of impulses. The decrease of L_{Aeq8} levels corresponding to 150- and 170-dB peak SPLs and the average decrease of the L_{Aeq8} and L_{eq8} (unweighted) levels are also given in Table I.

The decrease in L_{Aeq8} seen in Fig. 11 becomes larger as the A duration becomes shorter with the growth of peak SPL. This relation is similar to that obtained for the peak SPLs plotted in Fig. 3, except that the lines are less steep. Therefore, for A duration values above 0.5 ms, at peak SPLs below 163 dB, the decrease in the L_{Aeq8} under the earmuff is 2–4 dB more than the reduction of peak SPL. The opposite is true for short A durations of 0.3–0.5 ms, at peak SPLs above 165 dB, where the decrease in L_{Aeq8} is 2–5 dB less than the decrease in peak SPL. As was discussed earlier, this difference can be explained by the low-pass frequency characteristics of earmuff attenuation. High-level impulses, with short A duration, contain more energy in the high-frequency spectrum range, in which the energy is attenuated by the earmuff most efficiently. In contrast, most of the energy of impulses with long A duration falls into the low-frequency

range, in which the attenuation by the earmuff is less efficient. The low-pass frequency characteristics of earmuffs is also the reason for the lower—by about 10 dB—decrease in the unweighted equivalent level L_{eq8}, in comparison with the decrease in the A-weighted equivalent level L_{Aeq8} (see Table I). Due to the earmuff's less efficient attenuation at low frequencies there is also almost no change in L_{eq8} when the peak level of the outside impulse is increased. The values shown in Table I also indicate that the decrease in L_{Aeq8} is in slightly better agreement (about 2–3 dB) with the SNR value specified by the manufacturer than the decrease in peak SPL.

The hearing damage risk criterion based on A-weighted 8-h equivalent daily energy exposure level[20,25] and the 85-dB upper exposure limit classifies the impulses generated by the source used in the present study as generally safe, as for the 170-dB peak SPL the L_{Aeq8} equals 84 dB and decreases to about 70 dB for peak SPL of 150 dB (Fig. 11). According to the criterion based on L_{Aeq8}, an exposure to one impulse of 170-dB peak SPL and an exposure to 30 impulses of 150-dB peak SPL are both equivalent to the exposure allowed for an 8-h work day.

The attenuation of $L_{A\text{eq}8}$ by earmuffs shown in Fig. 11 indicates that the exposure under the earmuff is well below the danger levels, according to Dancer's criterion. At 170-dB peak SPL, for instance, the use of the least efficient earmuff (Sonico Standard, left panel) would theoretically make it possible to increase the number of exposures to 100 impulses per 8 h, whereas wearing the most efficient earmuff (Peltor H10, Bilsom 747, right panel) permits an exposure to more than 3000 impulses per 8 h. At 150-dB peak SPL, the average allowable number of exposures to impulses ranges from 1000 to 3000 for light and medium-size earmuffs (Fig. 11, left and middle panels), and equals about 4000 for heavy earmuffs (Fig. 11, right panel). This result demonstrates that the $L_{A\text{eq}8}$ criterion is the least stringent of all the criteria discussed earlier, which is in an agreement with the opinion expressed in the literature.[25] However, according to many opinions[11] the risk of inner-ear damage increases substantially when the peak SPL exceeds 140 dB, making the $L_{A\text{eq}8}$ criterion applicable only to lower levels.

VI. CONCLUSIONS

The main findings obtained in the present study are as follows:

(1) Transmission loss measurements of 32 different types of earmuffs showed that acoustic impulses with short A, C, or D durations are more effectively attenuated than longer impulses. On average, the attenuation of peak SPL is 10 dB less when the A duration is increased from 0.4 to 1.1 ms. This result is in agreement with published data obtained in field measurements for various kinds of shooting impulses.

(2) The SNR parameter only approximately represents the decrease in peak SPL. The decrease in peak SPL is predicted fairly well by the SNR value for A durations of about 0.45 ms.

(3) The impulse's rise time and A duration changes under the earmuff in a nonproportional manner. A rise time of $12-400$ μs and A duration of $0.4-1.1$ ms outside the earmuff increases to $0.3-1.5$ and $1.5-4.5$ ms, respectively, under the earmuff.

(4) Accounting for the use of hearing protectors by a simple level adjustment underestimates the risk of hearing damage assessed using impulse noise damage risk criteria. A meaningful estimate of hearing damage risk should also include the change in the A, C, or D durations of a signal under the hearing protector.

ACKNOWLEDGMENTS

[1] E. H. Berger, "Methods of measuring the attenuation of hearing protection devices," J. Acoust. Soc. Am. **79**, 1655–1687 (1986).

[2] S. Savolainen and K. M. T. Lehtomäki, "Hearing protection in acute acoustic trauma in Finnish conscripts," Scand. Audiol. **25**, 53–58 (1996).

[3] J. H. Patterson and D. L. Johnson, "The effects of exposure to intense free field impulse noise on humans wearing hearing protection: implication for new criteria," in *Proceedings of the 16th International Congress on Acoustics*, Seattle, Washington, 1998.

[4] G. Permentier, A. Dancer, K. Buck, G. Kronenberg, and C. Beck, "Artificial head (ATF) for evaluation of hearing protectors," Acust. Acta Acust. **86**, 847–852 (2000).

[5] A. W. Bronkhorst and G. F. Smoorenburg, "Assessment of hearing protector performance in impulsive noise: Update of research activities within the C-Funded IMPRO Project," in *Scientific Basis of Noise-induced Hearing Loss*, edited by A. Axelsson, H. M. Borchgrevink, R. P. Hamernik, P.-A. Hellstrom, D. Henderson, R. J. Salvi (Thieme, New York, 1993), pp. 339–348.

[6] G. F. Smoorenburg, "Assessment of hearing protector performance in impulsive noise. Final report," TNO Report TM-96-C042, Netherlands (1996).

[7] E. H. Berger, "Preferred methods for measuring hearing protector attenuation," in *Proceedings of Inter noise RIO 2005-Environmental noise control*, Rio de Janeiro, Brazil, 2005.

[8] M. Ylikoski, J. Pekkarinen, J. Starck, R. Pääkkönen, and J. Ylikoski, "Physical characteristics of gunfire impulse noise and attenuation by hearing protectors," Scand. Audiol. **24**, 3–11 (1995).

[9] A. Dancer, R. Franke, G. Parmentier, and K. Buck, "Hearing protector performance and NIHL in extreme environments: Actual performance of hearing protectors in impulse noise/nonlinear behaviour," in *Scientific Basis of Noise-induced Hearing Loss*, edited by A. Axelsson, H. M. Borchgrevink, R. P. HamernikP.-A. Hellstrom, D. Henderson, R. J. Salvi (Thieme, New York, 1993), pp. 321–338.

[10] J. O. Pekkarinen, J. P. Starck, and J. S. Ylikoski, "Hearing protection against high-level shooting impulses in relation to hearing damage risk criteria," J. Acoust. Soc. Am. **91**, 196–202 (1992).

[11] J. Ylikoski, J. Pekkarinen, and J. Starck, "The efficiency of earmuffs against impulse noise from firearms," Scand. Audiol. **16**, 85–88 (1987).

[12] E. Toppila and J. Starck, "The attenuation of hearing protectors against high-level shooting impulses," Arch. Acoust. **29**(2), 275–282 (2004).

[13] ANSI S12.42-1995, "Microphone-in-real-ear and acoustic test fixture methods for the measurement of insertion loss of circumaural hearing protection devices," (American National Standards Institute, New York, 1995).

[14] J. Starck, E. Toppila, H. Laitinen, G. Suvorov, V. Haritonov, and T. Grishina, "The attenuation of hearing protectors against high-level industrial impulse noise; comparison of predicted and *in situ* results," Appl. Acoust. **63**, 1–8 (2002).

[15] G. Parmentier and K. Buck, "Determination of the characteristics of earmuffs subjected to the action of impulse noise," in *Proceedings of the International Symposium on Effects of Impulse Noise on Hearing*, Malmo, Sweden, 1980.

[16] R. Pääkkönen, K. Lethomäki, J. Myllyniemi, E. Hämäläinen, and S. Savolainen, "Noise attenuation of hearing protectors in the human ear—a method description," Acust. Acta Acust. **86**, 477–480 (2000).

[17] O. Woxen and H. M. Borchgrevink, "Attenuation of hearing protectors at 85 dB SPL investigated by commercial 'insertion gain' method," in *Proceedings of the XXI Nordic Congress of Military Medicine*, Olso, 1990.

[18] ISO 4869-3, "Acoustics—Hearing protectors—Part 3: Simplified method for the measurement of insertion loss off ear-muff type protectors for quality inspection purposes," ISO/TR 4869-3 (1989).

[19] J. Pfretzchner, A. Moreno, and C. Colina, "The role of the physical factors involved in sound-attenuation characteristics of ear muffs," Appl. Acoust. **36**, 1–17 (1992).

[20] A. Dancer, K. Buck, G. Parmentier, and P. Hamery, "The specific problems of noise in military life," Scand. Audiol. **27**, 123–130 (1998).

[21] F. Pfander, H. Bongartz, H. Brinkmann, and H. Kietz, "Danger from auditory impairment from impulse noise: A comparative study of the CHABA damage-risk criteria and those of the Federal Republic of Germany," J. Acoust. Soc. Am. **67**, 628–633 (1980).

[22] F. Pfander, "Damage risk criteria with and without ear protection for impulse noise with high intensities regarding ear, larynx and lungs," in *Hearing and Hearing Prophylaxis*, Proceeding of the Olso International Symposium on the Effect of Noise on Hearing, Scand. Audiol. Suppl. **16**, 41–48 (1982).

[23] W. W. Dixon, "Proposed damage-risk criterion for impulse noise (gunfire)," Working Group 57, NAS-NRC Committee on Hearing, Bioacoustics, and Biomechanics, NAS-NRC Committee on Hearing, 1–10 (1968).

[24] A. Dancer, P. Grateau, A. Cabanis, T. Vaillant, and D. Lafont, "Delayed temporary threshold shift induced by impulse noises (weapon noises) in

men," Audiology **30**, 345–356 (1991).

[25]A. Dancer and R. Franke, "Hearing hazard from impulse noise: A comparative study of two classical criteria for weapon noises (Pfander criterion and Smoorenburg criterion) and the $L_{A\text{eq}8}$ method," Acust. Acta Acust. **3**, 539–547 (1995).

[26]M. R. Forrest, "The efficiency of hearing protection to impulse noise," in *Proceedings of the International Symposium on Effects of Impulse Noise on Hearing*, Malmo, Sweden, 1980.

[27]J. Zera and Z. Puslowski, "A foil blaster impulse noise source," in *Proceedings of Noise Control' 01*, Kielce, Poland, 2001.

[28]ISO 10843, "Acoustics – Methods for the description and physical measurement of single impulses or series of impulses" (Intl. Organization for Standardization, Geneva, 1997).

[29]ISO 4869-2, "Acoustics—Hearing protectors—Part 2: Estimation of effective A-weighted sound pressure levels when hearing protectors are worn," (Intl. Organization for Standardization, Geneva, 1994).

[30]EN 458, "Hearing protectors - Recommendations for selection, use, care and maintenance—Guidance document" (British Standards Institution, London, 2004).

[31]K. Buck, "Performance of hearing protectors in impulse noise," RTO Lecture Series 219, Report RTO-ENP-11 AC/323(HFM)TP/31, Res. and Tech. Org. (RTO) of NATO, Neuilly-Sur-Seine Cedex, France, 3-1–3-10 (2000).

[32]R. Pääkkönen, "Effects of cup, cushion, band force, foam lining and various design parameters on the attenuation of earmuffs," Noise Control Eng. J. **38**(2), 59–65 (1992).

[33]R. P. Hamernik and K. D. Hsueh, "Impulse noise: Some definitions, physical acoustics and other considerations," J. Acoust. Soc. Am. **90**, 189–196 (1991).

[34]G. F. Smoorenburg, "Damage risk criteria for impulse noise," in *New Perspectives on Noise*, edited by R. P. Hamernik, D. Henderson, and R. Salvi (Raven New York, 1982), pp. 471–490.

Optimal poroelastic layer sequencing for sound transmission loss maximization by topology optimization method

Joong Seok Lee, Eun Il Kim, and Yoon Young Kim[a]
National Creative Research Initiatives Multiscale Design Center, School of Mechanical and Aerospace Engineering, Seoul National University, Shinlim-Dong San 56-1, Kwanak-Gu, Seoul 151-742, Korea

Jung Soo Kim and Yeon June Kang
Advanced Automotive Research Center, School of Mechanical and Aerospace Engineering, Seoul National University, Shinlim-Dong San 56-1, Kwanak-Gu, Seoul 151-742, Korea

Optimal layer sequencing of a multilayered acoustical foam is solved to maximize its sound transmission loss. A foam consisting of air and poroelastic layers can be optimized when a limited amount of a poroelastic material is allowed. By formulating the sound transmission loss maximization problem as a one-dimensional topology optimization problem, optimal layer sequencing and thickness were systematically found for several single and ranges of frequencies. For optimization, the transmission losses of air and poroelastic layers were calculated by the transfer matrix derived from Biot's theory. By interpolating five intrinsic parameters among several poroelastic material parameters, distinct air-poroelastic layer distributions were obtained; no filtering or postprocessing was necessary. The optimized foam layouts by the proposed method were shown to differ depending on the frequency bands of interest.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2770541]

I. INTRODUCTION

This work addresses a new design method for maximizing the sound transmission loss of a one-dimensional multilayered acoustical foam by topology optimization. Multilayered acoustical foams consist of poroelastic material layers and air layers (air gaps). For efficient and accurate acoustical analysis required for foam design, a transfer matrix derived from Biot's theory[1] is used. The use of a transfer matrix greatly enhances computational efficiency. To find an optimal layer sequence consisting of an air layer and a certain poroelastic material by a gradient-based topology optimization formulation, material interpolation functions must be properly selected. Because there are several poroelastic material parameters, it is important to select a minimum set of material parameters for fast, reliable optimization convergence. The interpolation function should be appropriately selected to ensure air gap—poroelastic material distributions at the end of optimization iterations. Due to the lack of related research, the topology optimization for poroelastic acoustic foam design is a challenging subject.

One-dimensional multilayered acoustical foams composed of poroelastic materials and air gaps have been used in various applications to attenuate noise in automobile, aircraft, and construction industries. In many situations, multilayered acoustical foams can be analyzed by a simplified one-dimensional model because designers are mainly concerned with noise transmission through the multilayer rather than detailed two- or three-dimensional pressure distribu-

tions. Based on this observation, the present topology optimization to find optimal air and poroelastic layers sequencing employs a one-dimensional multilayered acoustical foam model. More specifically, one-dimensional sound transmission loss maximization of a two-material foam consisting of air and a poroelastic material will be formulated as a topology optimization problem. In this work, air implies a material forming an air layer (air gap) while air in pores of a poroelastic material will be referred to as the fluid phase of a poroelastic material.

When both a poroelastic material and air are used to form a multilayered acoustical foam, its acoustical performance is determined by the amount of the poroelastic material and air-poroelastic layer sequencing. If the amount of a poroelastic material in a multilayered acoustical foam is limited, it is difficult to determine the optimal distribution of air and a poroelastic material merely by trial and error. Moreover, because acoustical performance of a foam strongly depends on excitation frequencies, optimal foam configurations must be different depending on frequencies of interest. To overcome these difficulties, we will formulate the multilayered foam design problem as a topology optimization problem. No research directly related to the topology optimization of the foam is available, but it is important to review investigations on wave transmission through multilayered acoustical systems.

Thomson[2] was concerned with elastic wave transmission through a stratified periodic solid medium. Delany and Bazley[3] carried out the characteristic impedance analysis of porous materials. And Lauriks *et al.*[4] characterized a porous medium in laterally infinite plate-porous-plate systems for acoustic transmission analysis. Bolton and Green[5] studied

---

[a]Author to whom correspondence should be addressed; electronic mail: yykim@snu.ac.kr

the transmission of normally incident sound to laterally infinite double panel systems. Recently, Lee and Chen[6] reported acoustic transmission analysis of multilayered absorbers and Kobayashi et al.[7] studied wave transmission characteristics for periodic multilayers and fiber arrays having finite length. Optimal design issues were also discussed in these investigations, but no topology optimization studied.

To systematically find optimal poroelastic layer sequences maximizing sound transmission loss, we formulate the sequencing problem as a one-dimensional topology optimization problem. The advantage of using the topology optimization method is that no baseline design is required and thus truly optimal layer configurations can be determined. To carry out the optimization efficiently, acoustic analysis such as the calculation of sound transmission loss for a multilayered foam is performed by means of a transfer matrix derived by Bolton et al.[8] The transfer matrix[8,11] represents a relation between incoming waves to and outgoing waves from a multilayered acoustical foam. If air layer and a poroelastic material layer were treated separately, two sets of different governing equations would be needed, one for air layer based on the Helmholtz equation and the other for a poroelastic material layer based on Biot's theory. However, we use Biot's equations to model air layers as well as poroelastic foam layers for topology optimization formulation; to do so, poroelastic material properties are interpolated as functions of design variables to express both an air state and a poroelastic material state. Because air and a poroelastic material must be recovered in either limit of the design variable, no intermediate design value corresponding to an ambiguous material state should appear at the end of optimization iterations. Therefore, the success of the topology optimization heavily depends on the selection of proper material parameters to be interpolated and the form of interpolation functions—the interpolation issue is one of the most important parts in the present development. The detailed discussions on interpolation will be given in Sec. III. Note that because the topology optimization method does not require any initial configuration, no database constructed by repeat analyses for different configurations is necessary for this method.

By using the developed topology optimization formulation, optimal layer sequences of one-dimensional multilayered acoustical foams maximizing sound transmission loss were obtained under given mass usage constraint of a poroelastic material. The design optimization was conducted for several single frequencies and ranges of frequencies. For instance, the optimal layer sequence covering a wide frequency range between 1.0 and 5.0 kHz was obtained. No postprocessing or filtering was needed to obtain distinct air-poroelastic material layers at the end of the topology optimization process. The optimal results obtained by the proposed method appeared to be consistent with our physical intuition.

II. MATHEMATICAL DESCRIPTIONS

The first part of this section is concerned with a brief overview of Biot's theory,[1] which has been used to describe wave propagation in poroelastic materials. Then the transfer matrix approach used to evaluate sound transmission coeffi-

cient is presented. A more detailed account of the materials presented in this section may be found in Refs. 1 and 8–11.

A. Biot's theory

Biot[1] established a theory of elastic wave propagation for a coupled system consisting of a poroelastic solid saturated by a viscous fluid. The theory can be expressed in terms of the following:

$$N\nabla^2\mathbf{u} + \nabla[(A+N)e + Q\varepsilon] = \frac{\partial^2}{\partial t^2}(\rho_{11}\mathbf{u} + \rho_{12}\mathbf{U})$$

$$+ b\frac{\partial}{\partial t}(\mathbf{u} - \mathbf{U}), \tag{1a}$$

$$\nabla[Qe + R\varepsilon] = \frac{\partial^2}{\partial t^2}(\rho_{12}\mathbf{u} + \rho_{22}\mathbf{U}) - b\frac{\partial}{\partial t}(\mathbf{u} - \mathbf{U}), \tag{1b}$$

where $\mathbf{u}$ is vector-field solid displacement and $\mathbf{U}$ is vector-field fluid displacement in a poroelastic material. The symbol $e = \nabla\cdot\mathbf{u}$ denotes the volumetric strain of a solid part and $\varepsilon = \nabla\cdot\mathbf{U}$, the volumetric strain of a fluid part, respectively. The symbol ∇ represents the gradient operator. The elastic shear modulus is denoted by N and the first Lamé constant, by A. The coefficients Q and R express the coupling phenomena between the volume change of the solid phase of a poroelastic material and that of an interstitial fluid. The symbols ρ_{11}, ρ_{12}, and ρ_{22} denote mass coefficients accounting for the effects of nonuniform relative fluid flow through pores in a poroelastic material. The last term $b\cdot\partial(\mathbf{u}\text{-}\mathbf{U})/\partial t$ is the viscous coupling force where the parameter b represents the effect of viscous coupling. The coefficients (N, A, Q, R, ρ_{11}, etc.) appearing in Eq. (1) depend on several intrinsic parameters such as the modulus of elasticity,[1,8,11]

$$N = \frac{E_1}{2(1 + \nu)}, \tag{2a}$$

$$A = \frac{\nu E_1}{(1 + \nu)(1 - 2\nu)}, \tag{2b}$$

$$Q = (1 - h)E_2, \tag{2c}$$

$$R = hE_2, \tag{2d}$$

$$\rho_{11} = \rho_1 + \rho_a, \tag{2e}$$

$$\rho_{12} = -\rho_a, \tag{2f}$$

$$\rho_{22} = \rho_2 + \rho_a, \tag{2g}$$

where

$$\rho_1 = (1 - h)\rho_m, \tag{2h}$$

$$\rho_2 = h\rho_{\text{air}}, \tag{2i}$$

$$\rho_a = \rho_2(\varepsilon' - 1). \tag{2j}$$

In Eq. (2), E_1 and ν are the modulus of elasticity and Poisson's ratio of solid phase. Because of $E_1 = E_m(1+j\eta)$ (E_m: real-valued Young's modulus, η: loss factor), E_1 is complex valued. The symbols E_2 and h denote the bulk modulus of fluid in pores and the porosity of a poroelastic material. The symbols ρ_1 and ρ_2 denote mass of solid and mass of fluid per unit volume, respectively. In Eqs. (2h) and (2i), the mass densities of the solid phase and the fluid phase are denoted by ρ_m and ρ_{air}, respectively. The symbol ε' is the geometrical structure factor, which is associated with tortuosity. Note that E_2 is related to the bulk modulus of air (E_{air}) as[8]

$$E_2 = \frac{E_{air}}{1+f(s)}, \tag{3a}$$

where

$$f(s) = \frac{2(\gamma-1)}{s}\frac{J_1(s)}{J_0(s)}, \tag{3b}$$

$$s = \mathrm{Pr}^{1/2}\lambda_c\sqrt{-j}, \tag{3c}$$

$$\lambda_c = \frac{8\omega\rho_{air}\varepsilon'}{h\sigma}. \tag{3d}$$

In Eqs. (3c) and (3d), Pr is the Prandtl number and J_i is the Bessel function of order i. The symbols γ and σ denote the specific heat ratio of air and the steady-state macroscopic flow resistivity. (The symbol j denotes $\sqrt{-1}$.)

B. Transfer matrix

Bolton et al.[8] derived a transfer matrix for the analysis of sound transmission through multipanel structures lined with poroelastic materials. Because the efficient transfer matrix approach will be employed for poroelastic layer sequence optimization, the underlying equations are briefly given here.

Referring to Fig. 1(a), an acoustic wave is assumed to be normally incident into a layer of a poroelastic material. For one-dimensional analysis, wave transmission along the horizontal x axis will be considered. In this case, the general solution[8] in the panel may be found as

$$u_x = je^{jk_y y}\left[\frac{k_{1x}}{k_1^2}C_1 e^{-jk_{1x}x} - \frac{k_{1x}}{k_1^2}C_2 e^{jk_{1x}x} + \frac{k_{2x}}{k_2^2}C_3 e^{-jk_{2x}x}\right.$$
$$\left. - \frac{k_{2x}}{k_2^2}C_4 e^{jk_{2x}x}\right] + j\frac{k_y}{k_t^2}e^{-jk_y y}[C_5 e^{-jk_{tx}x} + C_6 e^{jk_{tx}x}], \tag{4a}$$

$$u_y = jk_y e^{-jk_y y}\left[\frac{C_1}{k_1^2}e^{-jk_{1x}x} + \frac{C_2}{k_1^2}e^{jk_{1x}x} + \frac{C_3}{k_2^2}e^{-jk_{2x}x}\right.$$
$$\left. + \frac{C_4}{k_2^2}e^{jk_{2x}x}\right] - j\frac{k_{tx}}{k_t^2}e^{-jk_y y}[C_5 e^{-jk_{tx}x} - C_6 e^{jk_{tx}x}], \tag{4b}$$

$$U_x = je^{-jk_y y}\left[b_1\frac{k_{1x}}{k_1^2}C_1 e^{-jk_{1x}x} - b_1\frac{k_{1x}}{k_1^2}C_2 e^{jk_{1x}x}\right.$$
$$\left. + b_2\frac{k_{2x}}{k_2^2}C_3 e^{-jk_{2x}x} - b_2\frac{k_{2x}}{k_2^2}C_4 e^{jk_{2x}x}\right]$$
$$+ jg\frac{k_y}{k_t^2}e^{-jk_y y}[C_5 e^{-jk_{tx}x} + C_6 e^{jk_{tx}x}], \tag{4c}$$

$$U_y = jk_y e^{-jk_y y}\left[b_1\frac{C_1}{k_1^2}e^{-jk_{1x}x} + b_1\frac{C_2}{k_1^2}e^{jk_{1x}x} + b_2\frac{C_3}{k_2^2}e^{-jk_{2x}x}\right.$$
$$\left. + b_2\frac{C_4}{k_2^2}e^{jk_{2x}x}\right] - jg\frac{k_{tx}}{k_t^2}e^{-jk_y y}[C_5 e^{-jk_{tx}x} - C_6 e^{jk_{tx}x}], \tag{4d}$$

$$\sigma_x = e^{-jk_y y}\left[\left(2N\frac{k_{1x}^2}{k_1^2} + A + b_1 Q\right)C_1 e^{-jk_{1x}x} + \left(2N\frac{k_{1x}^2}{k_1^2} + A\right.\right.$$
$$\left. + b_1 Q\right)C_2 e^{jk_{1x}x} + \left(2N\frac{k_{2x}^2}{k_2^2} + A + b_2 Q\right)C_3 e^{-jk_{2x}x}$$
$$+ \left(2N\frac{k_{2x}^2}{k_2^2} + A + b_2 Q\right)C_4 e^{jk_{2x}x} + 2N\frac{k_y k_{tx}}{k_t^2}(C_5 e^{-jk_{tx}x}$$
$$\left. - C_6 e^{jk_{tx}x})\right], \tag{4e}$$

$$s = e^{-jk_y y}[(Q + b_1 R)C_1 e^{-jk_{1x}x} + (Q + b_1 R)C_2 e^{jk_{1x}x} + (Q$$
$$+ b_2 R)C_3 e^{-jk_{2x}x} + (Q + b_2 R)C_4 e^{jk_{2x}x}], \tag{4f}$$

$$\tau_{xy} = e^{-jk_y y}N\left[\frac{2k_y k_{1x}}{k_1^2}(C_1 e^{-jk_{1x}x} - C_2 e^{jk_{1x}x})\right.$$
$$+ \frac{2k_y k_{2x}}{k_2^2}(C_3 e^{-jk_{2x}x} - C_4 e^{jk_{2x}x}) + \frac{(k_y^2 - k_{tx}^2)}{k_t^2}(C_5 e^{-jk_{tx}x}$$
$$\left. + C_6 e^{jk_{tx}x})\right], \tag{4g}$$

where u and U are solid and fluid displacements, respectively. The subscripts x and y denote the components in the x and y axis. The axial and shear stresses in the solid are denoted by σ_x and τ_{xy}. The stress in the fluid is denoted by s. The symbols k_1 and k_2 denote the wave numbers of two longitudinal waves, whereas k_t, denotes, the wave number of a rotational wave in the solid. Complex constants C_1 through C_4 represent the amplitudes of volumetric strains, while C_5 and C_6, represent the amplitudes of rotational strains. The symbols b_1, b_2, and g denote coupling coefficients.

When the vector notations $L = \{u_x\ u_y\ U_x\ \sigma_x\ s\ \tau_{xy}\}^T$ and $R = \{u_x\ u_y\ U_x\ \sigma_x\ s\ \tau_{xy}\}^T$ correspond to the physical quantities at the left- and right-hand sides of a poroelastic layer, Eq. (4) can be put into the following symbolic form:

$$S_L = e^{-jk_y y}[A]_{x=0}C, \quad S_R = e^{-jk_y y}[A]_{x=d}C, \tag{5a}$$

$$S_L = [A]_{x=0} \cdot [A]_{x=d}^{-1}S_R, \tag{5b}$$

with
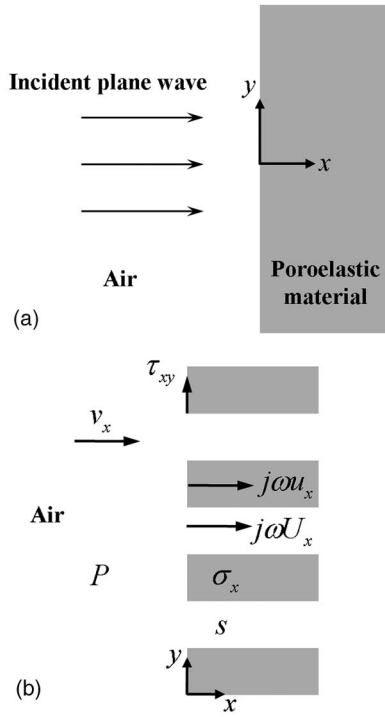
FIG. 1. Physical modeling of a poroelastic foam. (a) Normally incident plane wave model and (b) physical quantities appearing in the interface between a poroelastic foam (solid and fluid phases) and air.

$$C = \{C_1 \ C_2 \ C_3 \ C_4 \ C_5 \ C_6\}^T. \tag{5c}$$

In Eq. (5), $[A]$ is the matrix relating S_L and S_R to C. The width of a poroelastic layer is denoted by d.

The constants C_1 through C_6 in Eq. (5) can be determined when boundary conditions at the left- and right-hand sides of a poroelastic layer are given. If the condition of an open surface is used as illustrated in Fig. 1(b), the following hold:

$$-hP = s, \tag{6a}$$

$$-(1-h)P = \sigma_x, \tag{6b}$$

$$v_x = j\omega(1-h)u_x + j\omega h U_x, \tag{6c}$$

$$\tau_{xy} = 0, \tag{6d}$$

where P is pressure in the acoustic field exterior to the interface, v_x is the normal particle velocity of air at the interface, and ω is angular frequency.[9]

By imposing Eq. (6) on both interfaces of a poroelastic layer and substituting the results into Eq. (5), one can derive the relationship between pressure (P) and particle velocity (v_x) of the left-hand side and those of the right-hand side. The relationship is expressed as a 2 by 2 matrix $[\mathbf{TM}]$, which is known as a transfer matrix of a poroelastic material layer having open surfaces:[10,11]

$$\left\{ \begin{array}{c} P \\ v_x \end{array} \right\}_{\text{left}} = [\mathbf{TM}] \left\{ \begin{array}{c} P \\ v_x \end{array} \right\}_{\text{right}}. \tag{7}$$
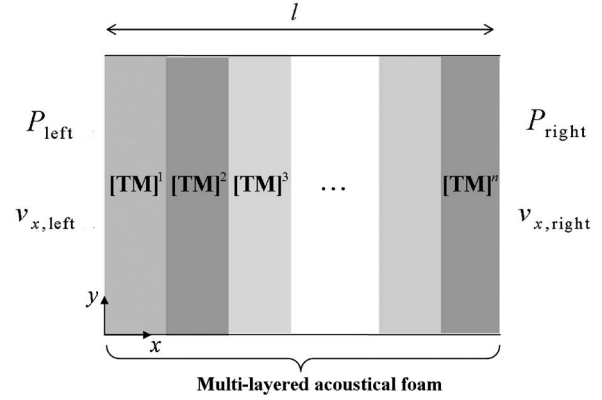
FIG. 2. Illustration of transfer matrices corresponding to individual layers of a multilayered acoustical foam system.

Because Eq. (7) applies only to a single layer under open surface interface conditions, the total transfer matrix for a multilayered acoustical foam (of length l) shown in Fig. 2 must be constructed by the multiplication of transfer matrices of all layers:

$$\left\{ \begin{array}{c} P \\ v_x \end{array} \right\}_{\text{left}} = [\mathbf{TM}]^{\text{Total}} \left\{ \begin{array}{c} P \\ v_x \end{array} \right\}_{\text{right}}, \tag{8a}$$

where

$$[\mathbf{TM}]^{\text{Total}} = [\mathbf{TM}]^1 [\mathbf{TM}]^2 \cdots [\mathbf{TM}]^n. \tag{8b}$$

Note that a multilayered foam in this work is made of several single layers that have open surface conditions [Eqs. (6)] at the left- and right-hand sides. Layers are treated as unbonded to each other. To facilitate the layer sequencing process by an optimization method, it is important to model a multilayered foam as a stack of unbonded layers satisfying conditions (6). Because Biot's compressive waves are effectively represented by P and v at the interface, the main physics of one-dimensional poroelastic layers can be captured. Indeed, actual numerical studies showed that the difference in transmission results from bonded multilayer and approximate unbonded multilayer analyses is insignificant. (The difference was less than 0.7 dB for an 8-cm-thick foam modeled with 80 layers.) Therefore, conditions (6) can be used to analyze each layer of a multilayered foam without hurting computational accuracy much.

Once $[\mathbf{TM}]^{\text{Total}}$ is computed, the reflection coefficient (R) and transmission coefficient (T) of a multilayered foam is straightforward. If the incident wave into the first layer on the left-hand side is a pressure wave of unit magnitude, the following equations can be used to define R and T in the first (left) and the last (right) layers of a multilayered foam:

$$P_{\text{left}} = 1 + R, \quad v_{x,\text{left}} = \frac{1}{\rho_{\text{air}} c}(1 - R), \tag{9a}$$

$$P_{\text{right}} = T e^{-jkl}, \quad v_{x,\text{right}} = \frac{1}{\rho_{\text{air}} c} T e^{-jkl}, \tag{9b}$$

where k is the wave number of the incident and reflected acoustic waves.

Using Eqs. (8a) and (9), the transmission coefficient T is given by

$$T = \frac{2e^{jkl}}{\mathrm{TM}_{11}^{\mathrm{Total}} + \left(\dfrac{1}{\rho_{\mathrm{air}}c}\right)\mathrm{TM}_{12}^{\mathrm{Total}} + \rho_{air}c \cdot \mathrm{TM}_{21}^{\mathrm{Total}} + \mathrm{TM}_{22}^{\mathrm{Total}}},$$

(10)

where $\mathrm{TM}_{ij}^{\mathrm{Total}}(i,j=1,2)$ are the components of the total transfer matrix $[\mathbf{TM}]^{\mathrm{Total}}$. In Eq. (10), $\rho_{\mathrm{air}}c$ denotes the normal specific acoustic impedance, where c is the phase velocity of acoustic waves. Because only normally incident plane waves are considered, Eq. (10) does not explicitly depend on an incident angle. Thus, U_y was eliminated from S_L and S_R of Eq. (5).

To analyze a multilayered acoustical foam consisting of a certain kind of poroelastic material layer and air layer, two transfer matrices, one for a poroelastic material and the other for the layer of air, may be separately constructed from different governing equations. However, if material properties of a poroelastic material are properly adjusted, the transfer matrix for an air layer can be computed from the transfer matrix of a poroelastic material. Because material property adjustment is directly related to the material interpolation scheme of the topology optimization, it will be fully discussed in Sec. III.

III. TOPOLOGY OPTIMIZATION FORMULATION

The following presents the main development for the topology optimization of one-dimensional multilayered acoustical foams. After the problem definition in topology optimization setting is given, material interpolation, crucial for solution convergence by a gradient-based optimization method, will be discussed extensively.

A. Topology optimization formulation

Topology optimization has been very successful in many structural design problems (see, e.g., Refs. 12 and 13). Though interest in the topology optimization of acoustic problems has recently increased,[14,15] no topology optimization formulation of poroelastic acoustical foams has been developed so far. In this section, we will formulate a topology optimization problem to find an optimal layer sequence of one-dimensional multilayered acoustical foams to maximize sound transmission loss. The sound transmission loss (TL) is defined as $\mathrm{TL}=10\log 1/|T|^2$, where T is the sound transmission coefficient defined in Eq. (9b).

Referring to Fig. 3, the topology optimization for optimal multilayered acoustical foams for some single frequencies or a range of frequencies can be stated as

$$\min_{\chi_e} F(\chi_e;f_l,f_u) = \min_{\chi_e}\sum_{f=f_l}^{f_u}|T(\chi_e;f)|$$

(11a)

subject to a mass constraint

$$\sum_{e=1}^{N_e}\chi_e \leq V_0.$$

(11b)
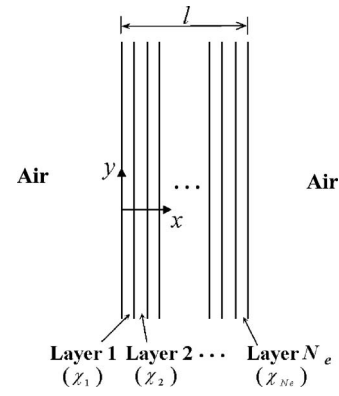
FIG. 3. Discretization of a design domain occupied by a multilayered foam. Each design variable $\chi_e(e=1,\ldots,N_e)$ controls the material state of the eth layer.

In Eq. (11a), F denotes an objective function expressed in density design variables χ_e ($1 \leq e \leq N_e$; N_e: number of discretized layers). The design variable χ_e that is assigned to each layer of the discretized foam in Fig. 3 can vary from 0 to 1. The poroelatic material properties will be interpolated as the functions of χ_e such that the states of air and a poroelastic material correspond to $\chi_e=0$ and $\chi_e=1$, respectively. Note that the thinner the layer thickness is, the higher the sequencing resolution can be.

To emphasize the dependence of the sound transmission coefficient T on frequency f, $T(\chi_e;f)$ is used in Eq. (11a). When a range of frequencies between the lower frequency f_l and the upper frequency f_u is considered, a simple discrete combination of $T(\chi_e;f)$ at some selected frequencies is used as the objective function F. Equation (11b) implies that the maximum volume usage of a poroelastic material is limited by V_0.

To solve the minimization problem, Eq. (11), by a gradient-based optimizer, the sensitivities of the objective function and the constraint equation with respect to the design variables must be computed. Though computation of $T(\chi_e;f)$ from Eq. (10) can be analytically derived, it is nearly impractical to derive the analytic expression for the sensitivity $\partial T/\partial \chi_e$ because $T(\chi_e;f)$ is a complicated function of χ_e. Thus, a direct finite difference method was used to calculate $\partial T/\partial \chi_e$ by using a small perturbation $\Delta=0.001$ of the design variable. Fortunately, the overall computational load by the finite difference is insignificant because the evaluation of $T(\chi_e;f)$ by the transfer matrix approach is extremely efficient. As an optimizer, the method of moving asymptotes[16] was used for all problems considered in this investigation.

B. Interpolation of poroelastic material properties

The following is concerned with a material property interpolation strategy. The interpolation should be able to represent the state of poroelastic material for $\chi_e=1$ and the state of air for $\chi_e=\chi_{e,\min}$ ($\chi_{e,\min}=0.01$ was used). At the same time, the interpolation functions should yield distinct air—poroelastic material layers at the end of optimization iterations because intermediate design variable statuses may be only artificial. Unfortunately, there are many parameters

TABLE I. Parameters of a poroelastic material used in the present investigation.

Material parameter	Value
Solid density (ρ_m)	552.5 kg/m³
In *vacuo* bulk young's modulus (E_m)	2.2×10^5 Pa
In *vacuo* loss factor (η)	0.98
Bulk Poisson's ratio (ν)	0.01
Flow resistivity (σ)	67 670 MKS Rayls/m
Porosity (h)	0.819
Structure factor (ε')	2.57

(a)

(b)

(c)

FIG. 4. The variation of TL values when the solid density (ρ_m) and/or the porosity (h) are interpolated. Interpolation of (a) ρ_m only, (b) h only, and (c) ρ_m and h.

characterizing material properties in Biot's theory. Therefore, it is important to work with the proper number of parameters for successful topology optimization. The main issue here is which material parameters should be chosen to achieve these goals. For all simulations in this section, the foam of length $l=8$ cm was controlled by a single design variable.

First, we begin the material interpolation with two fundamental parameters in Biot's theory; the solid density ρ_m and the porosity h. For instance, one may choose the following interpolation functions, commonly used in topology optimization (see, e.g., Ref. 12)

$$(\rho_m)_e = (\rho_m)_0 \, \chi_e^p, \tag{12}$$

$$h_e = (h_0 - 1)\chi_e^q + 1 \tag{13}$$

$$(0 < 0.01 = \chi_{e,\min} \leq \chi_e \leq 1).$$

In Eqs. (12) and (13) and throughout this work, the subscript 0 represents the nominal material values of the given poroelastic layer corresponding to $\chi_e = 1$. All material data are listed in Table I.

Figure 4 shows the TL values when the solid density and/or the porosity vary with Eqs. (12) and (13) (penalty exponents $p=1$ and $q=1$ were used). When $\chi_e = 1$, the TL value in Fig. 4 is equal to that of the given poroelastic material. The TL value decreases as χ_e decreases, but the TL behavior for $\chi_e = \chi_{e,\min}$ is quite off from the behavior of air; when χ_e becomes $\chi_{e,\min}(=0.01)$, the TL value should be sufficiently close to zero for all frequencies. (Because attenuation in air is not considered in this work, TL values for air layer should be zero.) When both ρ_m and h are interpolated, the TL values are improved slightly, but TL for $\chi_e = \chi_{e,\min}$ is still far from zero. Therefore, additional parameters must be interpolated to correctly simulate the state of air as the limit of a poroelastic material.

To look for additional parameters to be interpolated, physical phenomena occurring in poroelastic foams are examined. When an acoustic wave propagates through a poroelastic material, acoustic energy is dissipated due to the existence of the solid phase. Therefore, the coefficient b in Biot's theory representing viscous dissipation must be considered. In addition, the bulk modulus of fluid in pores (E_2) should be also considered because the complex-valuedness of E_2 implies acoustic energy dissipation.

To see the roles of b and E_2, the following interpolations are considered:
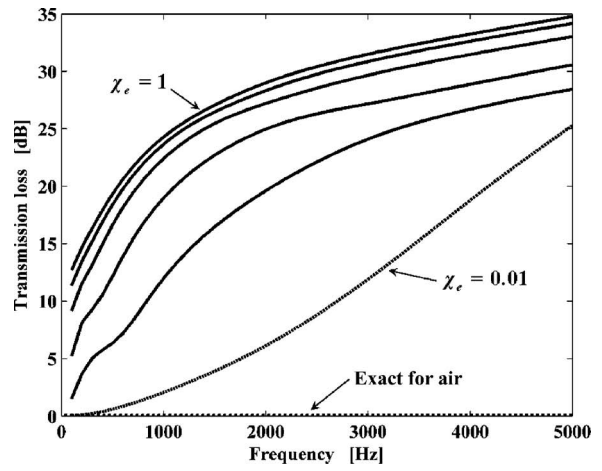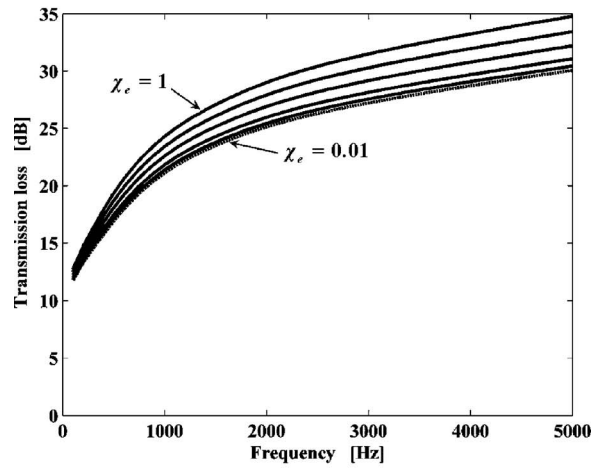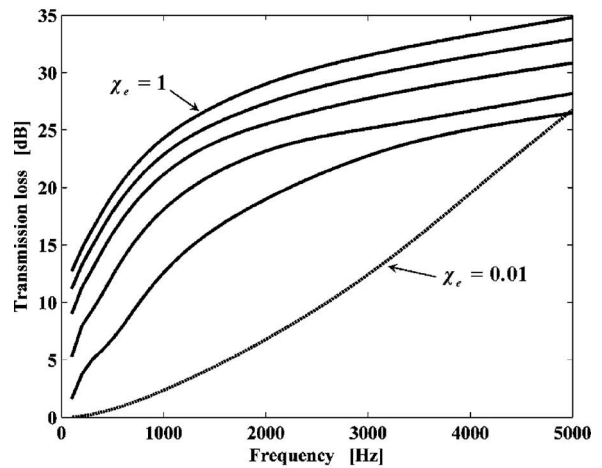
$$b_e = b_0 \, \chi_e^s, \tag{14}$$

$$(E_2)_e = \frac{E_0}{1 + f(s) \cdot \chi_e^t}. \tag{15}$$

The effect of the rational function in Eq. (15) will be discussed later, but in the meantime, we will mainly check how
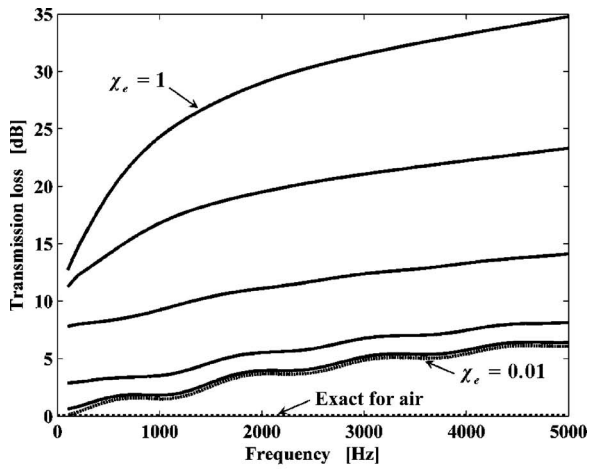
(a)
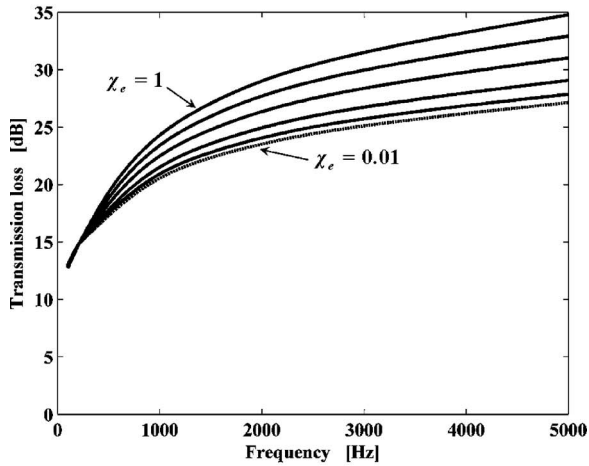
(b)
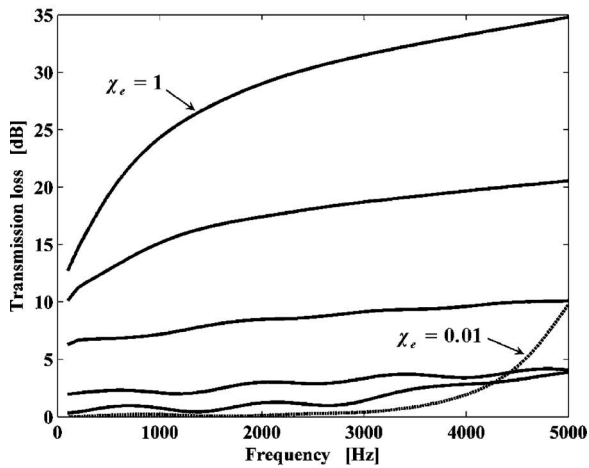
(c)

FIG. 5. The variation of TL values when the dissipation coefficient (b) and the bulk modulus of fluid in pores (E_2) are interpolated. (a) Interpolation of b only, (b) E_2 only, and (c) the simultaneous interpolation of ρ_m, h, b, and E_2.

closely the TL value can approach zero as b and E_2 decrease. Figures 5(a) and 5(b) ($s=2$, $t=1$ were used) plot the TL values for various values of χ_e when only b and E_2 are interpolated, respectively. In comparison with the results in Figs. 4(a) and 4(b), the controlling b is more effective in simulat-
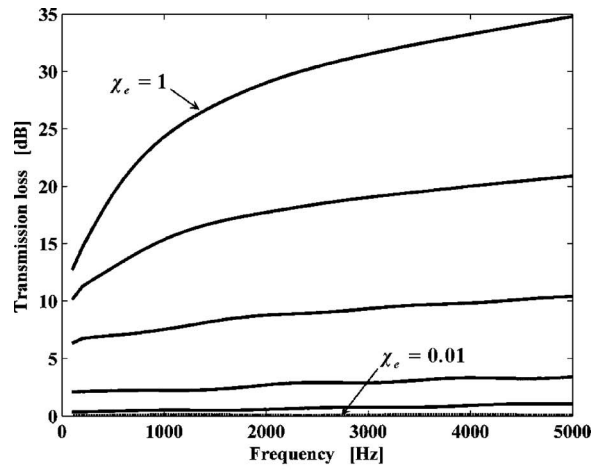
(a)
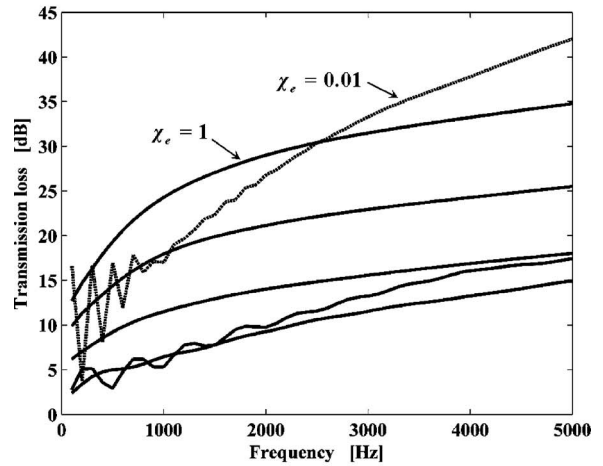
(b)

FIG. 6. The variation of TL values when five material parameters (ρ_m, h, b, E_2, and ε') are simultaneously interpolated where E_2 is interpolated by (a) Eq. (15) and (b) $(E_2)_e = E_{air}\chi'_e/[1+f(s)]$.

ing the TL behavior of air with a poroelastic material. However the results are not accurate.

On the other hand, Fig. 5(c) depicts the TL values when four parameters, ρ_m, h, b, and E_2, are simultaneously interpolated. The interpolated TL curve approaches the exact TL curve of air as χ_e approaches 0.01. However, it has erroneous behavior, especially in a region of high frequencies. This indicates that using four parameters is not still enough to represent the poroelastic state and the air state.

Finally, another parameter, the structure factor ε', is considered for interpolation,

$$\varepsilon'_e = \varepsilon'_0 \chi^r_e. \qquad (16)$$

Figure 6(a) shows the behavior of TL when five parameters (ρ_m, h, b, E_2, and ε') are simultaneously interpolated as the functions of χ_e [$r=0.25$ was used for Eq. (16)]. The air state having TL$=0$ is almost exactly simulated as the limit of the poroelastic material state with $\chi_e \approx 0$. Furthermore, the TL curves exhibit no oscillations for a wide range of frequencies in the present case.

The validity of the rational function of χ_e in Eq. (15) can be justified from physical reasoning. When χ_e approaches

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Lee *et al.*: Optimal poroelastic layer sequencing   2103

zero, the poroelastic material must simulate air. Therefore, E_2 must approach E_{air} as χ_e becomes zero. If one simply interpolates E_2 as $(E_2)_e = E_{\text{air}}\chi_e^f/[1+f(s)]$, a strange behavior shown in Fig. 6(b) will appear in the limit of $\chi_{e,\min}=0.01$.

Another important observation in material parameter interpolation is that the interpolation of the material parameters E_1 and ν (or N and A) of the solid phase is not used. In structural problems, these values (at least, E_1) must be interpolated to simulate both the solid and void states, but this is not the case for the present problem. In theory, the interpolation of E_1 [such as $(E_1)_e = (E_1)_0 \chi_e^\alpha$, α: interpolation exponent $\geqslant 1$] in addition to the above mentioned interpolations should work, but it was very difficult to find satisfactory interpolation function for N and A. To see why E_1 (or N, A) do not need to be interpolated, consider Biot's equations with the limiting values of the five parameters $(\rho_m, h, b, E_2, \text{ and } \varepsilon')\big|_{x_e\to 0}$ as determined by Eqs. (12)–(16). The limiting state should correspond to the state of air. It can be observed that in the limit, the Biot's equations in Eq. (1) are completely decoupled. Equation (1a) represents the time-independent governing equation solely for the solid state, while Eq. (1b) represents the time-dependent governing equation solely for the air state. Therefore, it is not necessary to interpolate E_1 and ν (or N and A).

IV. NUMERICAL CASE STUDIES

As numerical case studies, the topology optimization of one-dimensional multilayered acoustical foams having finite length of $l=8$ cm (see Fig. 3) was conducted. Because one-dimensional topology optimization problems are solved, the acoustical field is assumed not to vary in the y direction. In all cases considered, the incident plane waves were assumed to be normal to the left surface of the design domain. The design domain is discretized with 80 layers ($N_e=80$), each of which should be either a poroelastic material or air gap in the converged state. The material parameters of a typical poroelastic material in Table I are used for all simulations. Indeed, the parameters in Table I correspond to partially reticulated polyurethane foams. For all problems considered, the volume fraction ratio in Eq. (11b) is 70%.

A. Single frequency cases

The sound transmission loss maximization was solved for several single frequencies: $f=1.0$, 2.0, 3.0, 4.0, and 5.0 kHz. The application of the proposed topology optimization method yielded multilayered acoustical foams shown in Table II. The configurations and TL values of the optimized layer sequences are compared with those of a nominal layer configuration. The nominal foam was assumed to occupy 70% of the design domain by the poroelastic material without any intermediate air layer. The optimization iteration history for the target frequency $f=1.0$ kHz is plotted in Fig. 7(a). For all problems considered in this work, conversing results were obtained within 40–50 iterations. No postprocessing or filtering of the design values was used to obtain the results shown here. For the topology optimization, initial design values of 0.7 were assigned to all layers in the design domain for all problems.

TABLE II. Sound transmission loss (TL) values and foam shapes after foam layer sequencing optimization.

Target frequency	1.0 kHz	2.0 kHz	3.0 kHz	4.0 kHz	5.0 kHz
Nominal foam layer and TL values (dB)					
	17.7	21.1	22.8	24.1	25.0
Optimized foam layer and TL values (dB)					
	20.7	24.6	27.2	29.9	32.0
Performance increase (%)	17.1	16.5	19.0	24.3	28.0

From Table II, the following observation can be made. First, the topology optimization increased the sound transmission loss by 17%–28%, depending on frequencies. Second, symmetric layer configurations were obtained even without imposing explicit symmetry conditions. These symmetric layouts make sense because both sides of each opti-
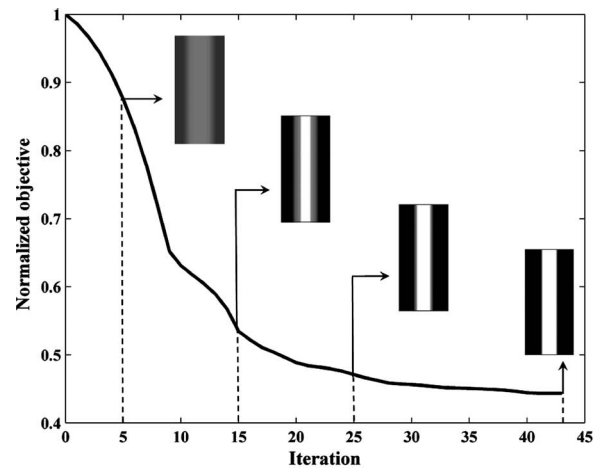
(a)
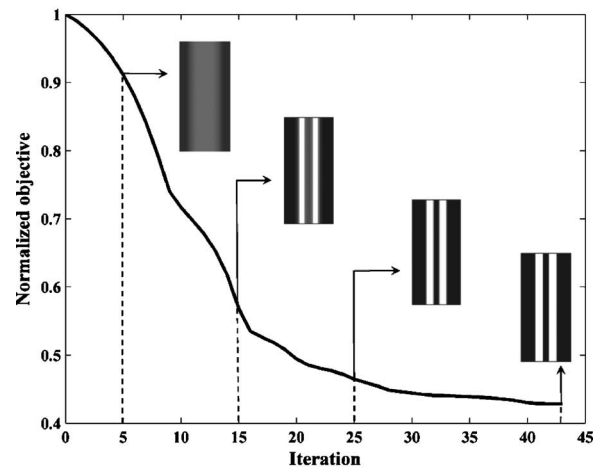
(b)

FIG. 7. Topology optimization history. (a) For a single frequency ($f=1.0$ kHz) and (b) for a frequency band (Band 1 with $f_l=1.0$ kHz, $f_u=2.0$ kHz).

mized foam interface with air. Third, air layers were located between layers of poroelastic material and the number of poroelastic layers increases as the target frequency increases while the thickness of each layer decreases. This is because wavelengths become shorter as frequencies become higher.

B. Narrow frequency band cases

In this case, the topology optimization method was used to find optimal poroelastic layer sequences for narrow frequency bands rather for single frequencies. The following three cases were considered:

Band 1: $f_l = 1.0$ kHz, $f_u = 2.0$ kHz,

Band 2: $f_l = 2.5$ kHz, $f_u = 3.5$ kHz,

Band 3: $f_l = 4.0$ kHz, $f_u = 5.0$ kHz.

In calculating the objective function in Eq. (11a), the summation over the frequencies at every 100 Hz from f_l to f_u was taken. The optimization history for Band 1 is shown in Fig. 7(b). As in single target frequencies, rapid solution convergence was achieved. No filtering or postprocessing was used. For all bands considered, the topology optimization improved TL by 14%–22% when the optimized foams are compared with the nominal foam.

It will be interesting to compare the results for single frequency cases and those for narrow frequency band cases in a single plot; see Fig. 8. As expected, the foams optimized for single target frequencies outperform those optimized for a frequency band. Interestingly, the optimal multilayered acoustical foam layer for a given band is similar to that obtained for the middle frequency of the band.

C. Wide frequency band case

As the last example, a wide frequency band ranging from $f_l = 1.0$ kHz to $f_u = 5.0$ kHz was considered. The summation in Eq. (11a) was taken at every 100 Hz. The sound transmission loss and the optimized multilayered sequence by the topology optimization are shown in Fig. 9. Even for this wide frequency band case, good solution convergence was observed and no postprocessing or filtering of the optimized result was necessary. The average sound transmission loss for the frequency range of the optimized foam increases up to 16.3% when the present foam is compared with the nominal foam.

V. CONCLUDING REMARKS

One-dimensional topology optimization formulation was investigated to design multilayered poroelastic acoustical foams. Optimal air-poroelastic multilayer sequences to maximize sound transmission loss were obtained by the proposed method for several single and ranges of frequencies. When the total amount of a poroelastic material is restricted, optimized foams turned out to consist of alternating layers of air and a poroelastic material of different thickness. Five intrinsic parameters appearing in Biot's theory (the solid density, porosity, viscous coupling effect, bulk modulus of fluid in pores, and structure factor of a poroelastic material) were
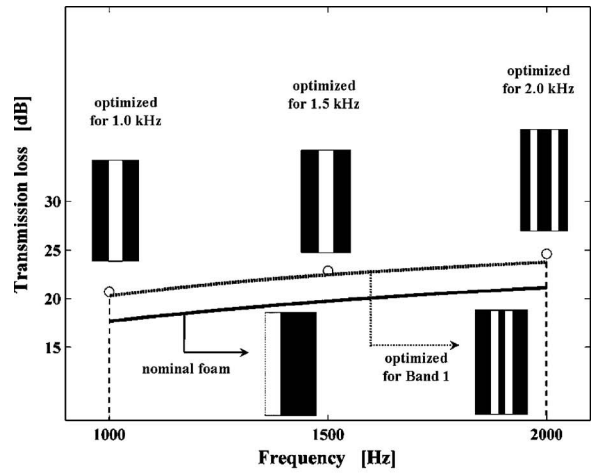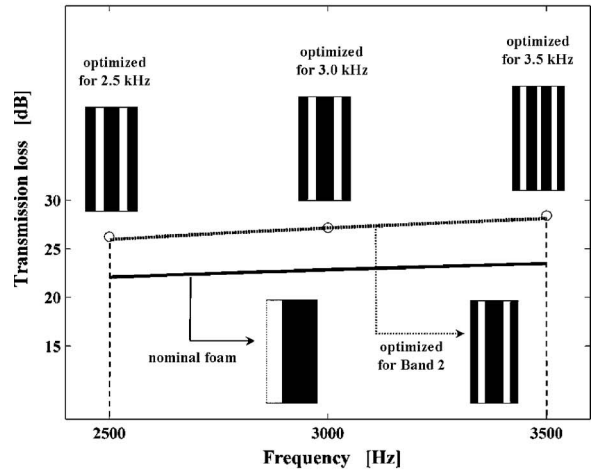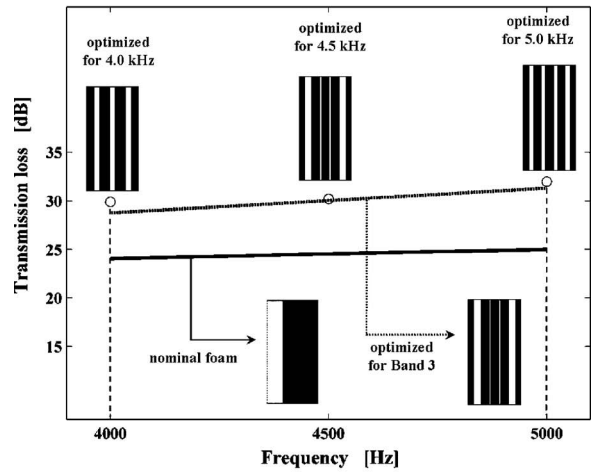
FIG. 8. Optimized results for three narrow frequency bands. (a) For Band 1 ($f_l = 1.0$ kHz, $f_u = 2.0$ kHz), (b) for Band 2 ($f_l = 2.5$ kHz, $f_u = 3.5$ kHz), and (c) for Band 3 ($f_l = 4.0$ kHz, $f_u = 5.0$ kHz). The TL values of the optimized foam are compared with those of the nominal foam. The circles denote the TL values of the foams optimized for single frequencies.

interpolated to represent all states between a poroelastic material and air. When these parameters were interpolated, distinct air-poroelastic material layers were obtained. No postprocessing or filtering was necessary. The interpolation of
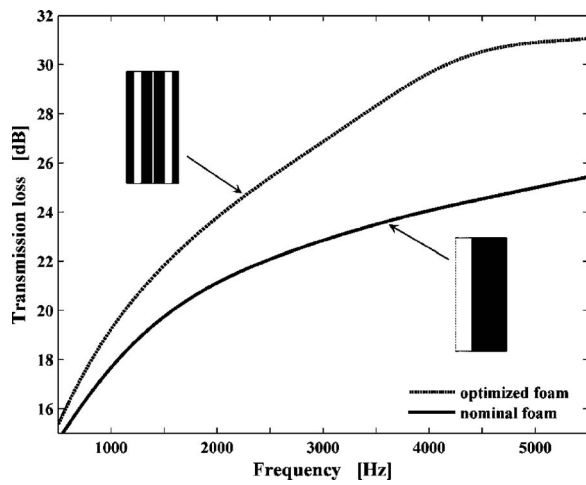
FIG. 9. The sound transmission loss of the foam optimized for a wide frequency band ($f_l = 1.0$ kHz, $f_u = 5.0$ kHz).

other parameters, such as the Lamé constant of the solid phase, was not needed. A transfer matrix derived from Biot's theory was used to calculate sound transmission loss. Numerical studies showed that the optimized foam layouts by the proposed topology optimization method outperform nominal foams for all frequencies considered.

ACKNOWLEDGMENTS

[1]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid," J. Acoust. Soc. Am. **28**, 168–191 (1956).

[2]W. T. Thomson, "Transmission of elastic waves through a stratified solid medium," J. Appl. Phys. **21**, 89–93 (1950).

[3]M. E. Delany and E. N. Bazley, "Acoustic properties of fibrous absorbent material," Appl. Acoust. **3**, 105–116 (1970).

[4]W. Lauriks, P. Mees, and J. F. Allard, "The acoustic transmission through layered system," J. Sound Vib. **155**, 125–132 (1992).

[5]J. S. Bolton and E. R. Green, "Normal incidence sound transmission through double panel system lined with relatively stiff, partially reticulated polyurethane foam," Appl. Acoust. **39**, 23–51 (1993).

[6]F. C. Lee and W. E. Chen, "Acoustic transmission analysis of multilayer absorbers," J. Sound Vib. **248**, 621–634 (2001).

[7]F. Kobayashi, S. Biwa, and N. Ohno, "Wave transmission characteristics in periodic media of finite length: Multilayers and fiber arrays," Int. J. Solids Struct. **41**, 7361–7375 (2004).

[8]J. S. Bolton, N.-M. Shiau, and Y. J. Kang, "Sound transmission through multi-panel structures lined with elastic porous materials," J. Sound Vib. **191**, 317–347 (1996).

[9]Y. J. Kang and J. S. Bolton, "Finite element modeling of isotropic elastic porous materials coupled with acoustical finite elements," J. Acoust. Soc. Am. **98**, 635–643 (1995).

[10]Y. J. Kang, "Studies of sound absorption by and transmission through layers of elastic noise control foams: Finite element modeling and effects of anisotropy," Ph.D. dissertation, School of Mechanical Engineering, Purdue University, West Lafayette, IN, 1994.

[11]J. F. Allard, *Propagation of Sound in Porous Media: Modeling Sound Absorbing Materials* (Elsevier Science, New York, 1993).

[12]M. P. Bendsøe and O. Sigmund, *Topology Optimization: Theory, Methods and Applications* (Springer, Germany, 2003).

[13]M. P. Bendsøe and N. Kikuchi, "Generating optimal topologies in structural design using a homogenization method," Comput. Methods Appl. Mech. Eng. **71**, 197–224 (1988).

[14]M. B. Dühring, "Topology optimization for acoustic problems," *IUTAM Symposium on Topological Design Optimization of Structures, Machines and Materials* (Springer, Netherlands, 2006), Vol. **137**, pp. 375–385.

[15]G. H. Yoon, J. S. Jensen, and O. Sigmund, "Topology optimization of acoustic-structure interaction problems using a mixed finite element formulation," Int. J. Numer. Methods Eng. **70**, 1049–1075 (2007).

[16]K. Svanberg, "The method of moving asymptotes: A new model for structural optimization," Int. J. Numer. Methods Eng. **24**, 359–373 (1987).

An experimental study of the acoustic impedance characteristics of human hair

Bradley E. Treeby,[a] Jie Pan, and Roshun M. Paurobally
Centre for Acoustics, Dynamics and Vibration, School of Mechanical Engineering, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

Previous analytical and empirical studies of the human auditory system have shown that the cues used for localization are modified by the inclusion of nonrigid scattering surfaces (clothing, hair etc). This paper presents an investigation into the acoustic impedance properties of human hair. The legitimacy of a locally reactive surface assumption is investigated, and an appropriate boundary condition is formulated to account for the physiological composition of a human head with hair. This utilizes an equivalent impedance parameter to allow the scattering boundary to be defined at a reference plane coincident with the inner rigid surface of the head. Experimental examination of a representative synthetic hair material at oblique incidence is used to show that a locally reactive surface assumption is legitimate. Additional experimental analysis of a simple scattering problem illustrates that the equivalent impedance must be used in favor of the traditional surface impedance to yield physically correct pressure magnitudes. The equivalent acoustic impedance properties of a representative range of human hair samples are discussed, including trends with sample thickness, fiber diameter, bulk density, and mass. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2773946]

I. INTRODUCTION

Below 3–4 kHz, the acoustic scattering characteristics of the human head match well with those of simple geometric shapes. As a result, sphere scattering models are commonly used to account for the principle features of human auditory localization cues over these frequencies. The traditional use of such models assumes that the head surface is completely rigid in nature.[1,2] However, the surface impedance of the head affects both the binaural localization cues,[3] and the head-related transfer function (HRTF).[4–6] Given the appropriate impedance data, recent mathematical treatments of the sphere scattering problem facilitate an anthropometric investigation into the contribution of hair to the auditory percept.[7] However, the acoustic properties of human hair are not well documented. While normal absorption coefficients for a small number of hair samples at arbitrary bulk densities have been presented,[8] this is not sufficient to extrapolate the required trends or variability in impedance values representative of typical human subjects. Other investigations have been interested in the characteristics only at higher frequencies, for example in response to ultrasound.[9] Additionally, no substantial discussion or investigation has been presented to address the validity of a locally reactive assumption for a covering of human hair. Measurements of the impedance of other fibrous materials at oblique incidence show that such materials can exhibit considerable properties of extended reaction, particularly at low frequencies.[10,11]

This paper investigates the legitimacy of a locally reactive surface assumption and formulates a boundary condition

appropriate to examine the acoustic characteristics of a human head with hair. This is validated through impedance measurement at oblique incidence and simple sphere scattering experiments. The acoustic impedance characteristics of a representative range of human hair samples are then investigated. Changes in impedance with sample thickness, fiber diameter, bulk density, and mass are discussed.

The physical properties and distribution of human hair (terminal scalp hair) typically vary with both ethnicity and hair color. The mean diameter of individual Caucasian hair fibers is around 72 μm with an ellipticity factor (ratio of maximum to minimum diameter) of 1.5.[12] Asian hair is generally thicker and rounder (77 μm and 1.28) and African hair thinner and flatter (66 μm and 1.84), although there is considerable variability within the groups, particularly for Caucasians.[12,13] Additional deviations are also evident along individual hair fiber lengths.[14] On average the human scalp contains between 175 and 300 hair fibers per square centimeter,[15,16] with a fiber density of approximately 1320 −1340 kg/m^3.[15,17] These parameters allow a realistic range for the bulk density of human hair to be approximated. If the hair is assumed to stand perpendicular to the scalp surface, the bulk porosity is in the range of 98.5–99.5% and the bulk density 10−20 kg/m^3. This alignment is realistic for short straight hair styled upwards. As the vertical hair is slanted back towards the scalp tangent, the hairs stack and the bulk density increases. Taking a parallelepiped segment of hair, as the hair is slanted the volume is scaled with angle (cos θ). Assuming that the segment retains a realistic finite volume, if the hair is sloped 75° from the scalp normal the bulk density range is increased to 40−80 kg/m^3 (94–97% bulk porosity). While the alignment of hair is more complex than this, and there are a multitude of fiber, distribution, and stylistic varia-

---

[a]Author to whom correspondence should be addressed. Electronic mail: treebs@mech.uwa.edu.au

tions that will change the overall bulk density, this provides a pragmatic approximation of the range. In comparison, sheep wool (also a largely keratin fiber) has a fiber diameter of $20-30$ μm, a fiber density of 1320 kg/m^3, and bulk densities in the range of $30-40$ kg/m^3.[15,18,19]

Modeling the acoustical properties of fibrous materials has received considerable attention over the last several decades, particularly in relation to the mineral fibers utilized for sound absorption. These models are generally either derived from empirical power law relations,[20–22] or detailed acoustic knowledge of the material microstructure.[23,24] The ability of these models to accurately account for the acoustic properties of materials is reliant on the match between the fiber alignment evident in the material structure, and the geometry and interactions assumed in the model formulation (frequently parallel fibers). In relation to the fibrous materials that are generally of interest to acousticians, human hair has relatively unique bulk structural composition. The hairs are anchored to the scalp in a uniform patterned arrangement, and then depending on their length either stand at least partially upright (short hair) or stack towards the rear or sides of the head depending on hairstyle (long hair). Material models that can account for the acoustic characteristics of isotropic or layered anisotropic fine-fiber materials such as mineral wools cannot necessarily be extrapolated to account for the properties of hair. In addition to variations in bulk structural composition, the diameter of individual hair fibers is significantly larger than many acoustic materials previously investigated. Consequently, the current work does not attempt to evaluate the suitability of fiber models for predicting the acoustic properties of human hair. However, the material parameters and experimental results presented facilitate such an analysis should it be of interest to future researchers.

II. HUMAN HAIR AS A LOCALLY REACTIVE BOUNDARY

A. Use of equivalent impedance

The solution of a scattering problem in an unbounded medium is reliant on the knowledge of the boundary conditions that exist between the scattering bodies and their surrounds. These boundary conditions are typically expressed in terms of normal acoustic impedance, the complex ratio of surface pressure to the velocity normal to the surface.[25] The velocity is produced either by the flow of the surrounding fluid through porous surface openings, or by motion of the surface boundary itself.[26] The impedance is dependent on the detailed properties and structure of the material which are assumed to be time invariant. If a material is acoustically "rigid," the surface impedance is much greater than that of the surrounding fluid. Consequently, the normal velocity at the boundary surface is approximately zero. Conversely, if a material is acoustically "soft" (or pressure released), the surface impedance is much less than that of the surrounding fluid and the total pressure on the surface is approximately zero. Physically, the real component of the impedance is associated with a damping force and corresponds to a net dissipation of energy. The imaginary part embodies the stiffness

and inertia of the surface and dictates the lag of the surface displacement behind the forcing pressure oscillation.

The analytic investigation of the contribution of hair to the auditory percept requires knowledge of these boundary characteristics, and their trends with representative modifications in hair diameter, thickness, and bulk density. Of particular interest to the current problem is whether the hair covering can be considered to be locally reactive. This requires that the motion of the surface at any position is dependent only on the incident acoustic pressure at that position, and is impartial to the pressure distribution over the remaining surface. The angle of refraction through the material is thus close to zero (the tangential component of the velocity is much less than the normal component) and the normal acoustic impedance is independent of the angle of wave incidence.[27] Note the reflection coefficient (and hence absorption coefficient) retains a characteristic dependence on the incidence angle, even for a locally reactive material.[25]

Human hair as a sparse acoustic material in isolation cannot legitimately be assumed to be locally reactive. It has a high porosity and low bulk density and consequently the fluid velocity in nonradial directions cannot be reasonably neglected. However, human hair in context has a relatively unique bulk composition. The hairs are anchored to the scalp in a uniform patterned arrangement and form only a thin layer over the acoustically rigid head. Of the surface boundary formulations utilized for reflection and scattering problems, two are available for consideration. First, if the surface can be considered locally reactive, the relationship between the pressure and radial velocity at any location on the surface is entirely determined by the normal acoustic impedance at that position. The solution for the scattered wave can then be formulated using knowledge of these surface characteristics considering only the exterior sound field. Alternatively, if the scatterer is constructed from a homogenous material or fluid that is extensively reactive, the solution for the scattered wave can be formulated using the density and compressibility of the interior material. Morse and Ingard[25] give a thorough treatment of both of these problems for spherical scatterers. For a penetrable sphere with inhomogeneous material parameters that remain spherically symmetric (i.e., the density and compressibility are only a function of radial distance), a similar integral formulation can also be formed.[28]

Given that the composition of interest approximates to a rigid sphere with a thin hemispherical covering of hair, the problem appears to correspond more closely to the assumptions of a locally reactive boundary than those of a penetrable sphere. While this may be a logical selection given the available boundary conditions, the assumption of a locally reactive surface does raise several other interesting questions. First, the traditional boundary formulation requires the surface impedance of the absorbent material. For a sparse fibrous material like hair, the location of this *surface reference plane* is not well defined or even necessarily immobile. Second, if a nonuniform impedance distribution is assumed, all impedance values must then be defined relative to this same surface reference plane [as shown in Fig. 1(a)]. Considering that the hair layer is atop the rigid head (in this case a sphere), this requires the impedance of the rigid sur-
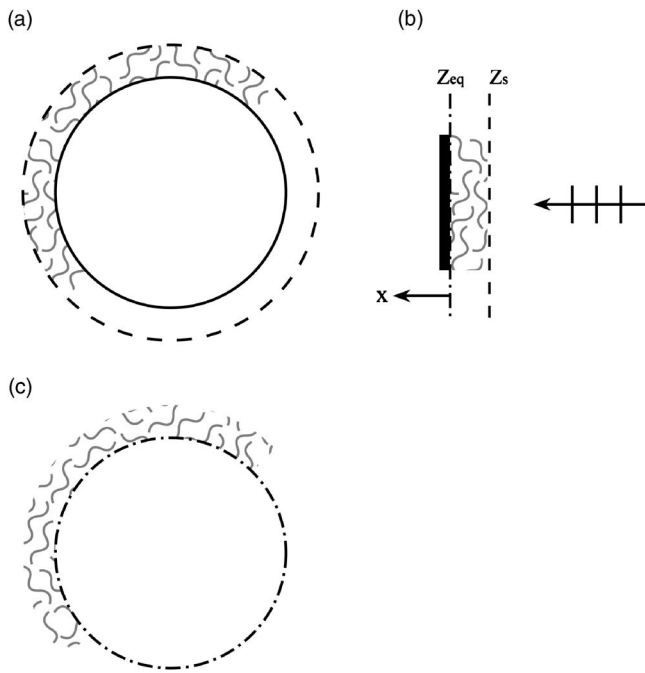
FIG. 1. (a) The traditional impedance formulation which requires the impedance of the rigid surface at an external reference plane. (b) The relationship between the equivalent and traditional impedance surface reference planes. (c) Reformulated scattering boundary condition using the equivalent impedance defined at the inner surface.

face to be defined at a reference plane exterior to the physical surface. For higher dimensional problems, the use of the normal surface impedance is no longer adequate to describe the characteristics of the material. This is because it is not legitimate to neglect the particle velocity in tangential directions. To correctly define the impedance would thus require three orthogonal components. Finally, as the formulation only accounts for wave motion exterior to the scattering surface, it is not valid to then examine the required pressure characteristics on the interior rigid surface which are obviously "inside" the defined surface reference plane.

Rather than reformulate the problem using a more complex composite boundary or to consider the impedance in three dimension, it is more convenient to define an *equivalent impedance* of the fibrous absorbent covering. This will be defined at a surface reference plane coincident with the interior rigid boundary, and formulated to adequately account for the absorption characteristics of the layer for all angles of incidence. Using the one-dimensional normal reflection problem for plane wave incidence (i.e., the impedance tube problem), the equivalent impedance Z_{eq} is computed by translating the traditional surface impedance Z_s to the new surface reference plane [see Fig. 1(b)]. Assuming a harmonic time component of the form $e^{-i\omega t}$, for a material of thickness ℓ this relationship is given by

$$\zeta_{eq} = \frac{(\zeta_s + 1) + (\zeta_s - 1)e^{-ik_0 2\ell}}{(\zeta_s + 1) - (\zeta_s - 1)e^{-ik_0 2\ell}}. \tag{1}$$

Here k_0 is the wave number in the propagation medium external to the material, and the specific acoustic impedance ζ is related to the normal acoustic impedance Z by the characteristic impedance $\rho_0 c_0$ of the external propagation medium

$$\zeta = \frac{Z}{\rho_0 c_0}. \tag{2}$$

Although the surface reference plane is defined coincident with the interior rigid boundary, the equivalent impedance defined here does not represent the impedance of this rigid surface in isolation. Rather, it continues to encapsulate the acoustic impedance properties of the material layer backed by the rigid boundary, the difference being in the nonconventional reference plane in which the impedance is defined. For the one-dimensional problem, it is straightforward to show that Z_s and Z_{eq} yield identical reflection (and absorption) coefficients. The solution for the time harmonic pressure (in one dimension) is thus independent of the reference plane in which the surface impedance is defined. It is important to note that the equivalent impedance values utilized throughout this paper can easily be converted to traditional surface impedance values using Eq. (1) and the thickness of the sample holder utilized for the impedance measurement. Likewise, any relevant surface impedance data that exist in the literature can be converted to equivalent impedance if the sample thickness is known.

As the equivalent impedance encapsulates only the properties of the material for normal incidence, its use for higher-dimensional problems is only valid provided that its value does not significantly change with incidence angle. This is examined experimentally in Sec. II B. For a sphere with a nonuniform boundary, the definition of equivalent impedance now allows the scattering boundary to be defined on the interior surface of the sphere as shown in Fig. 1(c). The use of this equivalent impedance parameter in conjunction with a locally reactive surface boundary offers computational advantages over more complex boundary conditions. The pressure and velocity on the boundary can now be related simply by the impedance of the surface material, rather than the maintenance of a pressure and velocity continuity across the boundary considering the wave motion inside the scattering object.

B. Measurement of acoustic impedance at oblique incidence

To examine the validity of a locally reactive surface assumption, the acoustic properties of human hair at oblique incidence were experimentally examined using a representative synthetic fibrous material. The properties of this material are shown in Table I (sample SH 01), and a large scale sample is visible in Fig. 2. The material was constructed with individual fibers attached at one end to a thin fabric backing with an approximate bulk density of 30 kg/m³. The material was initially selected as it had similar fiber diameter and bulk density characteristics to human hair, and then later verified to have comparable impedance properties. Reference measurements were made of the material using the two microphone impedance tube technique following the experimental procedure outlined by the ISO 10534-2 standard.[29] The impedance tube utilized had a circular internal cross section 60 mm in diameter, with a removable pipe section containing a movable rigid piston to allow sample placement. The working frequency range of the tube was 375–3000 Hz. Samples

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Treeby *et al.*: Acoustic characteristics of human hair   2109

TABLE I. The color, texture, treatment (N – natural, C – colored), curvature (s – straight, w – wavy, c – curly), diameter, and length of the human hair (HH) and synthetic hair (SH) samples.

Sample	Color	Texture	Treatment	Curvature	Mean diameter [μm]	Mean length [mm]
HH 01	Light brown	Thick	N	w	104	60
HH 02	Blonde	Medium	C	c	64	90
HH 03	Dark brown	Medium	C	s	77	60
HH 04	Dark blonde	Thick	N	s	75	40
HH 05	Light brown	Medium	N	c	80	50
HH 06	Light blonde	Fine	N	w	80	80
HH 07	Dark blonde	Medium	N	w	80	140
HH 08	Dark blonde	Fine	N	c	75	120
HH 09	Medium blonde	Medium	N	c	89	150
HH 10	Dark blonde	Medium	C	s	95	160
HH 11	Dark blonde	Medium	C	w	74	140
HH 12	Dark brown	Thick	N	c	85	70
SH 01	Medium brown	Fine	N	s	55	50

were held within a 40-mm-deep sample holder with a frontal termination made from an open wire grid. Two other smaller holders (20 and 10 mm in depth) were also constructed for additional tests. While the material had hair fibers up to 60 mm in length, the natural arrangement and pile of the bulk material meant that the samples were not significantly deformed when placed in the holder. The microphone separation and termination distances were calibrated by repeating multiple no sample tests with and without each of the sample holders (removing and replacing the microphones between each test). The distances were then calculated using the position of the first pole and first zero of the inter-microphone frequency response function and the speed of sound in the tube, averaged across the appropriate tests. The thin grid covering of the sample holders was shown to have a negligible effect on the measured parameters by repeating tests both with and without the grid present. The resulting surface impedance of the synthetic hair material measured in the impedance tube is shown in Fig. 3(a).

The measurement of normal acoustic impedance at oblique incidence has been of historical interest and a variety of test methodologies have previously been discussed.[10,30,31] The development of two microphone *in situ* measurement techniques has allowed this measurement to be simplified

significantly, and mathematical formulations using both plane and spherical waves have been proposed.[11,32,33] The
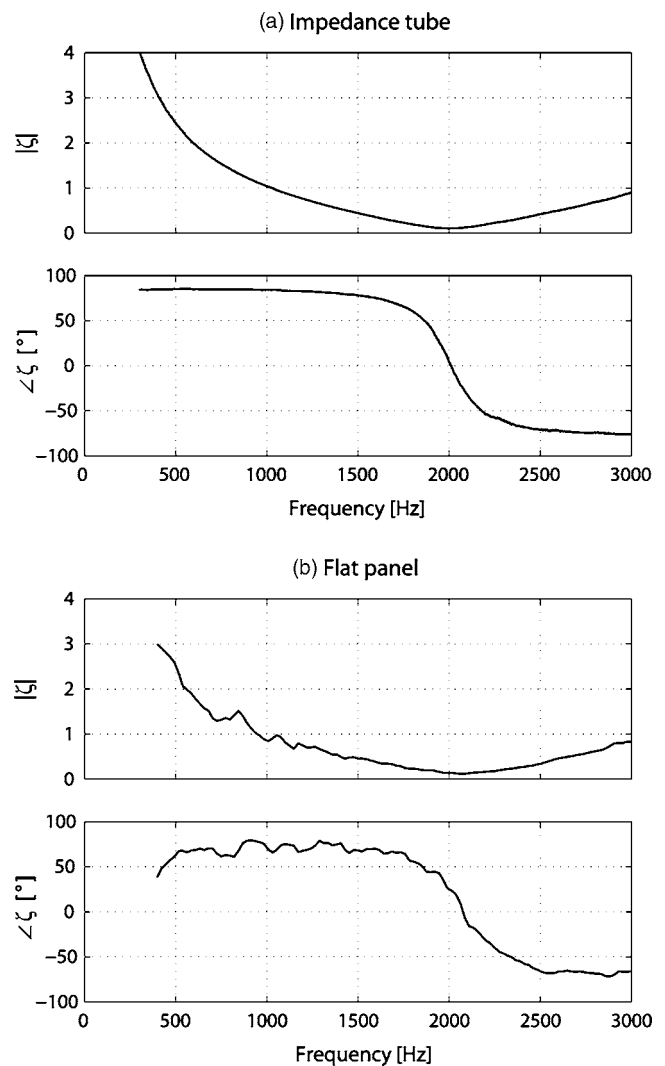
FIG. 3. Comparison of the specific normal surface impedance of the synthetic hair material for a random fiber alignment and a sample thickness of 40 mm. (a) Results from impedance tube tests. (b) Results from flat panel tests in free field using a microphone spacing of 15 mm and a surface offset of 63 mm.

FIG. 2. (Color online) Experimental setup for the measurement of normal acoustic impedance at oblique incidence using a spherical sound source. The synthetic hair material is shown with a random fiber alignment.

experiments presented here follow this two-microphone methodology assuming a spherical incident wave. Figure 2 shows the experimental setup with a large sample of the synthetic hair attached to a $1.2 \times 1.2 \times 0.025$ m panel made from high density fiber board. The wooden material was shown to be approximately rigid over the frequency range of interest (absorption coefficient $\alpha < 0.08$). Both random (hair gently swept against pile direction) and parallel (hair swept with pile direction) fiber alignments were tested. For each test, two 1/4 in. microphones (BSWA Tech MPA416) were held perpendicular to the panel, with either a 15 or a 50 mm separation (see Li and Hodgson[33] for a diagram of the standard experimental configuration). For the two fiber alignments, the closest microphone was positioned within 10 mm of the effective frontal surface (63 mm from the rigid backing for the random fiber alignment, 32 mm for the parallel). This was to minimize errors associated with the assumption of a constant source angle relative to material surface and the microphones.[33] A source separation of 0.4 m was used and incidence angles from 90° (normal incidence) to 40° were tested. For each angle of incidence, impulse response measurements were obtained using maximum length sequences (MLS) produced by the Brüel & Kjær DIRAC software. A horn driver fitted with a 0.2 m cylindrical attachment (19 mm internal diameter) was used to provide a spherical incident wave. A MLS length of $2^{14} - 1$ (the shortest available sequence length) with ten averages and a sampling frequency of 96 kHz was used. The impulse responses were shortened to 2^{12} samples, windowed using a cosine-tapered (Tukey) window with a 25% taper ratio, and converted to the frequency domain using a 2^{12} point fast Fourier transform. To compensate for any phase mismatch between the microphones, all tests were repeated with the microphone positions reversed.

Comparative trends for the surface impedance at normal incidence (from 400 Hz) are shown in Fig. 3(b). The results shown correspond to a random hair alignment and a microphone spacing of 15 mm. To allow direct comparison with the impedance tube measurement, this assumes a surface reference plane that is offset 40 mm from the rigid backing (equal to the depth of the impedance tube sample holder used). While the results from the free-field flat panel test contain significantly more noise, the two measurements are essentially identical. This mutually validates both experimental procedures. The noise in the free-field panel tests, particularly at low frequencies, is attributable to the frequency limitations of the sound source used, the finite size of the material sample, and the imperfect nature of the anechoic test environment. The exact impedance values for both test methodologies are strongly dependent on the selection of the location of the surface reference plane. For the free-field panel tests this was extremely difficult to estimate. The corresponding changes in the real and imaginary components of the normal surface impedance with incidence angle are shown in Fig. 4. Again this assumes a surface reference plane that is offset 40 mm from the rigid backing. Contrary to the discussion by Li and Hodgson,[33] the results from either microphone separation yielded nearly identical results over the entire frequency range of interest. For the same surface ref-
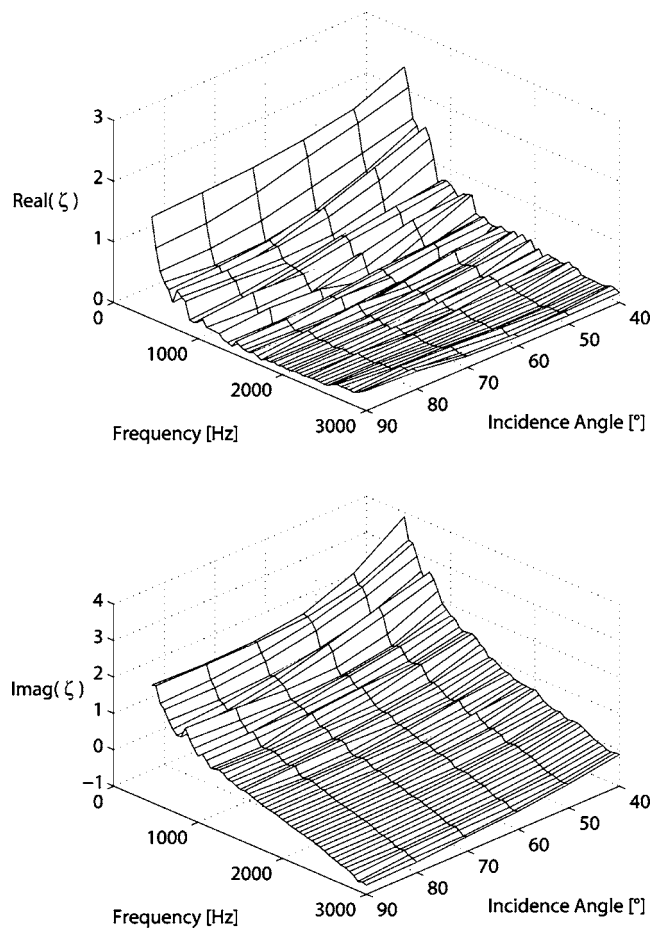
FIG. 4. Changes in the specific normal surface impedance of the synthetic hair material with incidence angle (real and imaginary components) for a random fiber alignment, a sample thickness of 40 mm, a microphone spacing of 15 mm, and a surface offset of 63 mm. Here 90° corresponds to normal incidence.

erence plane, the alignment of the fiber also had little effect, although a diminutive resonance was observable over a small frequency range for the parallel alignment tested. This resonance is due to the coupling between the motion of the air and the fibrous material.[34] As the bulk density is increased this becomes more significant, however, as tested the impedance properties are not strongly dependent on this alignment.

As shown in Fig. 4, at low frequencies there is a noticeable dependence of the normal impedance on incidence angle, and the material cannot be considered entirely locally reacting. As frequency increases, the changes with incident angle are reduced. This is consistent with other oblique incident tests of fibrous materials,[10,11,33] although a definitive acoustic explanation does not always accompany such results. For fibrous and porous materials, the acoustic impedance modalities necessitate a slightly different phenomenological explanation than other textbook acoustic interfaces. Such a description is given in detail by Ingard.[27] At very low frequencies, the physical structure of the material itself is able to vibrate and the acoustic impedance properties are dependent on the material's stiffness. The excitation and coupling of a limited number of structural wave modes within the material cause it to be extensively reactive (the excitation of the surface at one position will induce motion at other

locations). Similarly, at very high frequencies the air within the porous cavities will support wave motion similar to that in free space, and the surface impedance will again be extensively reactive. Between these extremities, the material frame is approximately rigid, and the friction force dominates both inertial or stiffness characteristics. In this circumstance, the wave motion in the material degenerates into a diffusive process, and the interior phase velocity is much smaller than that in the external propagation medium.[27] Consequently, the angle of refraction through the material is close to zero, and the normal acoustic impedance does not show a strong dependence on incidence angle.[23]

Even if the phase velocity in the material is not substantially below that in the external propagation medium (i.e., the refraction angle within the material is not necessarily close to zero), for a thin porous layer with a rigid backing, the material may nonetheless be approximately locally reacting. If the stiffness and inertia of the material do not facilitate additional wave motion, the area of *extended influence* (i.e., the region over which the pressure oscillation at any position has influence) can still be considered to be relatively small. In this situation, the normal impedance will again not show a strong dependence on incidence angle (away from grazing incidence). For the synthetic hair material, it is likely that both a reduced internal phase velocity and the small thickness of the layer contribute to the diminished observable variance in the normal acoustic impedance with incidence angle over the higher frequencies shown in Fig. 4.

The impedance properties of the material layer at higher frequencies cannot be intrinsically inferred from those discussed at lower frequencies. Without access to detailed material properties and adequate material models, it is difficult to estimate exactly how the material will behave, particularly at frequencies significantly above the measured range. This is due to the multiplicity of acoustic characteristics that porous layers may exhibit.[27] Frequency and other limitations currently exist in both free-field and impedance tube measurement techniques, and consequently published results rarely exceed 5 or 6 kHz. While this range is sufficient for the current problem (sphere scattering models of the human head), further classification may be required to describe the acoustic impedance properties of human hair over the complete audible frequency range.

As the surface reference plane is translated towards the rigid backing (where the equivalent impedance is defined), the overall impedance magnitude is increased. Figure 5 illustrates the corresponding equivalent impedance of the synthetic hair material. Any remaining incongruity with incidence angle at low frequencies is now of less relevance, as the impedance magnitude over this range dictates a surface that is approximately rigid. Considering the frequency range of interest for scattering models used by binaural synthesis (typically below 3 kHz), a locally reactive surface boundary can now be considered a legitimate assumption. Although not formally tested, it is intuitively expected that higher density hair samples would illustrate similar characteristics. Conversely, for samples with a substantially increased material thickness (a large afro, for example), a locally reactive equivalent impedance parameter may no longer adequately
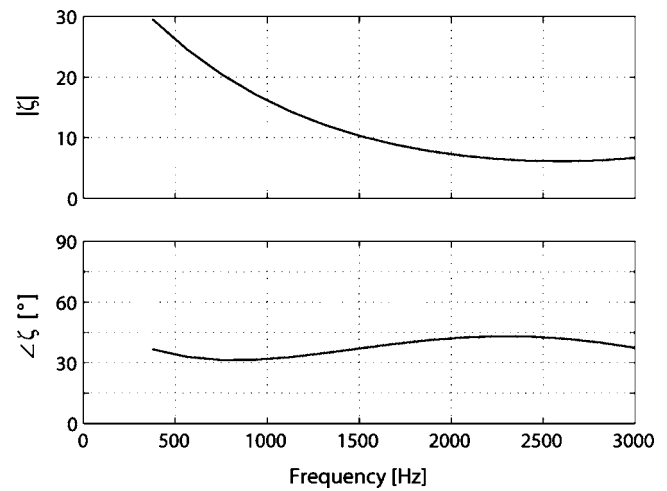
FIG. 5. Equivalent impedance of the synthetic hair material calculated using Eq. (1) and the surface impedance measured by the impedance tube shown in Fig. 3.

capture the acoustic properties of the layer, as a small area of extended influence no longer inherently exists.

C. Experimental validation of the use of equivalent impedance

To illustrate the legitimacy of using equivalent impedance for higher dimensional problems, the scattering of sound by a sphere with a uniformly distributed surface boundary will be briefly examined. The mathematical formulation of this problem is reviewed in many texts and is not repeated here, e.g., Ref. 3. Given a 0.124 m radius rigid sphere uniformly covered in the synthetic hair material, the pressure exterior to the surface can be calculated using the normal acoustic impedance. Figure 6 illustrates the resultant pressure magnitude at 1000 Hz along the anterior median axis adjacent to the sphere. The solid line corresponds to use of the equivalent impedance (shown in Fig. 5), and the dashed line the impedance defined at the traditional surface reference plane (shown in Fig. 3). To allow direct comparison, the results using the latter assume an enlarged sphere radius equal to the material thickness. Clearly the definition of the surface reference plane influences the analytical result.

To validate which impedance surface reference plane yields the physically correct pressure values, the problem was examined experimentally within an anechoic chamber. A 0.124 m radius wooden sphere supported by a thin steel rod was uniformly covered by the synthetic hair material using a thin double-sided tape. The covering was tailored circumspectly to maintain the overall distribution of the hair (with a bulk density of approximately 30 kg/m^3), and so that it fitted neatly over the sphere surface without any significant deformation. Seven 1/2 in. microphones (BSWA Tech MA211) were positioned in an evenly spaced array along the anterior median axis as shown in the upper panel of Fig. 6. The pressure perturbation due to the presence of the sphere was then obtained by measuring the frequency response between the microphone array and the excitation source (white noise from a Brüel & Kjær HP1001 unidirectional sound source) both with and without the sphere. The source was positioned

Treeby *et al.*: Acoustic characteristics of human hair
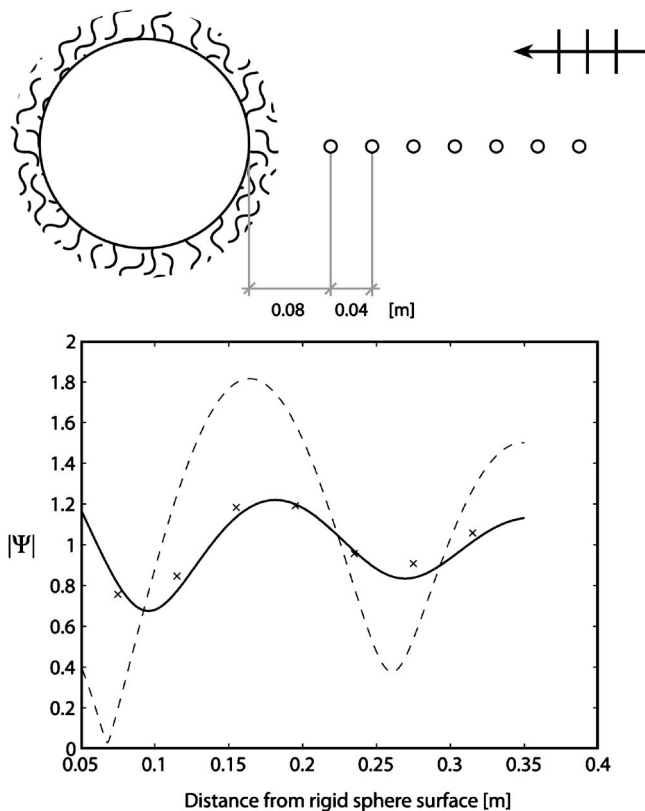
FIG. 6. The pressure magnitude $|\Psi|$ along the anterior median axis of a 0.124 m radius rigid sphere uniformly covered by a synthetic hair material, where Ψ is the spatially dependent component of the complex pressure. The upper panel illustrates the experimental setup showing the position of the evenly spaced array of seven microphones (circles). The incident wave approaches from the right and is assumed to be plane. The lower panel illustrates the experimental pressure magnitude at 1000 Hz (crosses), along with comparative analytic values using the equivalent impedance (solid line) and the traditional surface impedance (dotted line). The impedance characteristics used are shown in Figs. 5 and 3, respectively.

approximately 2.5 m from the sphere to allow plane wave propagation to develop. The measurements were captured and processed using the National Instruments LabVIEW software, and an eight-channel data acquisition card (National Instruments 4472B). Again both random (hair gently swept against pile direction) and parallel (hair swept with pile direction) fiber alignments were tested.

The pressure magnitudes at the seven microphone positions are illustrated by the crosses in Fig. 6. The results shown correspond to the test using a random hair alignment and match closely with the analytical results using the equivalent impedance. As with the oblique incidence tests the alignment of the hair had little effect on the measured pressure magnitudes. At other frequencies the experiment yielded analogous results. However, it should be noted that at significantly higher frequencies the spatial resolution of the microphones was not high enough to give an accurate comparison. For the current work it is sufficient to conclude that the use of the equivalent impedance yields legitimate analytical results. Experimental analysis of the sphere surface pressure (not discussed here) yielded the same supposition.

A similar scattering problem was examined by Cook and Chrzanowski[35] to determine the absorption coefficient of a cattle-hair felt. A series of rigid spheres covered with a layer

of the hair felt were placed within an anechoic chamber. Comparative results derived from a locally reactive sphere scattering model using the surface impedance of the hair felt deviated substantially from experimental results. The differences were attributed to the inadequacy of the boundary condition to account for the properties of the felt. Experimental results presented by Brungart[36] for the pressure adjacent to a rigid sphere uniformly covered with kapok (simulating hair) also differed substantially from analytical results. The analytical results were derived using measured values of the surface admittance assuming a locally reactive surface. The simulations using the kapok admittance (defined at the traditional surface reference plane) more closely resembled the characteristics of an infinitely soft sphere than those of a rigid sphere. This is consistent with the results presented in Fig. 6. The differences in Brungart's results were again attributed to the inadequacy of the measured surface impedance to account for the properties of the material.

III. NORMAL ACOUSTIC IMPEDANCE PROPERTIES OF HUMAN HAIR

A. Validation of measurement technique

As the formulation and validation of an appropriate impedance boundary is now complete, it is useful to examine the extent of equivalent impedance values experienced by a representative range of human hair samples. As it is not possible to easily measure hair impedance *in situ* from human subjects, the acoustic properties will again be examined using an impedance tube. The experimental results discussed in the preceding sections suggest that the impedance is not strongly dependent on the relative arrangement of the hair fibers. However, for these experiments the individual fibers remained attached to a fabric backing. Samples of human hair collected from hair cuttings are no longer constrained in such a manner. The consequence of this constraint on measured impedance values was examined using three samples of the synthetic hair material each attached directly to a rigid wooden backing with thin double-sided tape. The impedance properties of the material samples were measured, then the hair fibers cut from the backings and remeasured under the same conditions using a variety of fiber alignments. The mean (solid line) and range (dashed lines) of the measured impedance values are shown in Fig. 7(a). The variability is well within the expected range for multiple sample tests, and there is clearly not a strong dependence on the attachment of the material to the rear surface.

To examine the impedance trends of human hair, 12 samples of varying hair type and length were collected from cuttings from a salon. The properties of these samples are shown in Table I. The mean diameter was measured using a profile projector (± 1 μm) and three hair fibers selected at random from each sample (each fiber was measured at two different locations). The color, texture, treatment, and curvature were classified by the salon from which the samples were collected. The equivalent impedance variability across all samples for a density of 40 kg/m^3 and a sample thickness of 20 mm is shown in Fig. 7(b). Again, the results across all samples are reasonably consistent, particularly at high fre-
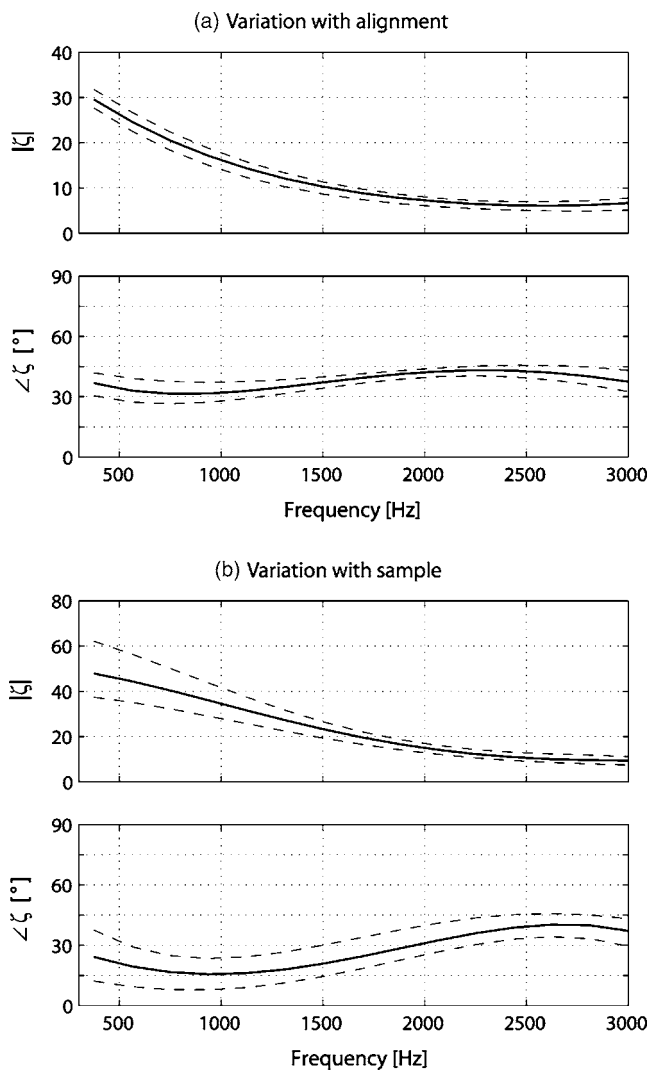
FIG. 7. (a) Variation in the equivalent impedance of the synthetic hair material with sample alignment. (b) Variation across all human hair samples for a sample density of 40 kg/m^3 and a sample holder thickness of 20 mm.

quencies where the material is absorbent. There is not a strong correlation between the spread of results and the properties of the individual hair samples. While it is generally expected that flow resistance and hence impedance will change with both bulk density and fiber diameter, for larger diameters this variability is considerably reduced.[37] The differences seen may be partly attributed to the difficultly in arranging the fibers homogenously within the sample holder, particularly for longer samples. It is also interesting to note that there is significantly less observable variance in the traditional surface impedance computed from the same material tests.

At this point it is useful to briefly discuss the consequences of the choice of the harmonic time component (which is assumed here to be of the form $e^{-i\omega t}$). While this selection is arbitrary, care must be taken that it is rigorously maintained. Analytical problem formulations, impedance measurements, and any Fourier transforms or signal analysis techniques utilized must all exhibit the same dependence. Many presentations of impedance data do not state the form of the time component used. Likewise, the ISO two micro-

phone impedance tube standard[29] does not explicitly state which form it assumes in its analytical formulation. This is significant as this selection affects the physical interpretation of the impedance results. For a harmonic time component of the form $e^{-i\omega t}$, a positive reactance (positive impedance phase angle) corresponds to a stiffness-like response, and a negative reactance to a mass-like response. If $e^{i\omega t}$ is assumed, the impedance becomes its complex conjugate and the opposite is true.

B. Impedance trends

Considering that the impedance properties of the hair samples do not show a strong correlation with individual sample parameters, it is sufficient to investigate changes in equivalent impedance with sample density and thickness using a subset of these. Samples HH 01 through HH 04 were tested using the impedance technique previously described for a variety of bulk densities using three sample holders (10, 20, and 40 mm in depth). The corresponding impedance and absorption coefficient results are shown in Fig. 8. The plots of absorption coefficient (right panels) also illustrate the variability across the four tested samples (dashed lines). Figure 8(a) illustrates the variation with sample density using a constant sample holder thickness of 20 mm. These plots show the variance between people with a high natural bulk density of hair (i.e., a high number of hairs per cm^2) and those with a low bulk density. Obviously there is no theoretical minimum to this trend and it is not completely uncommon to see people without any scalp hair. In this situation a uniformly rigid boundary is an adequate approximation. The variance with sample thickness for a constant sample density of 40 kg/m^3 is shown in Fig. 8(b). These plots illustrate the effect of growing shorter hair outwards at an approximately constant bulk density. For people with longer hair, if the hair is not styled upright the individual fibers stack parallel and the impedance properties over the scalp remain reasonably impervious to hair growth. Figure 8(c) again shows the variance with sample thickness, this time for a constant sample mass. This is equal to the mass of the hair sample at 40 kg/m^3 held within the 20 mm sample holder. These plots illustrate the effect of compressing and separating the same quantity of hair fibers.

Examining the trends shown within Fig. 8, it is clear that the impedance properties scale with both sample density for constant thickness, and sample thickness for constant density. Increasing the bulk density reduces the impedance magnitude (and increases the absorption coefficient) while the impedance phase angle remains reasonably constant. Increasing the sample thickness also reduces the impedance magnitude, and additionally scales the impedance phase angle. The changes with thickness for constant mass are considerably less, but a decrease in impedance magnitude (and increase in absorption) is again seen with an increase in sample thickness. For all modifications the equivalent impedance retains a stiffness-like reactance, and the impedance phase angle remains largely within the range of 10–50°.

It is important to note that for the one-dimensional problem the choice of surface reference plane does not alter the

Treeby *et al.*: Acoustic characteristics of human hair
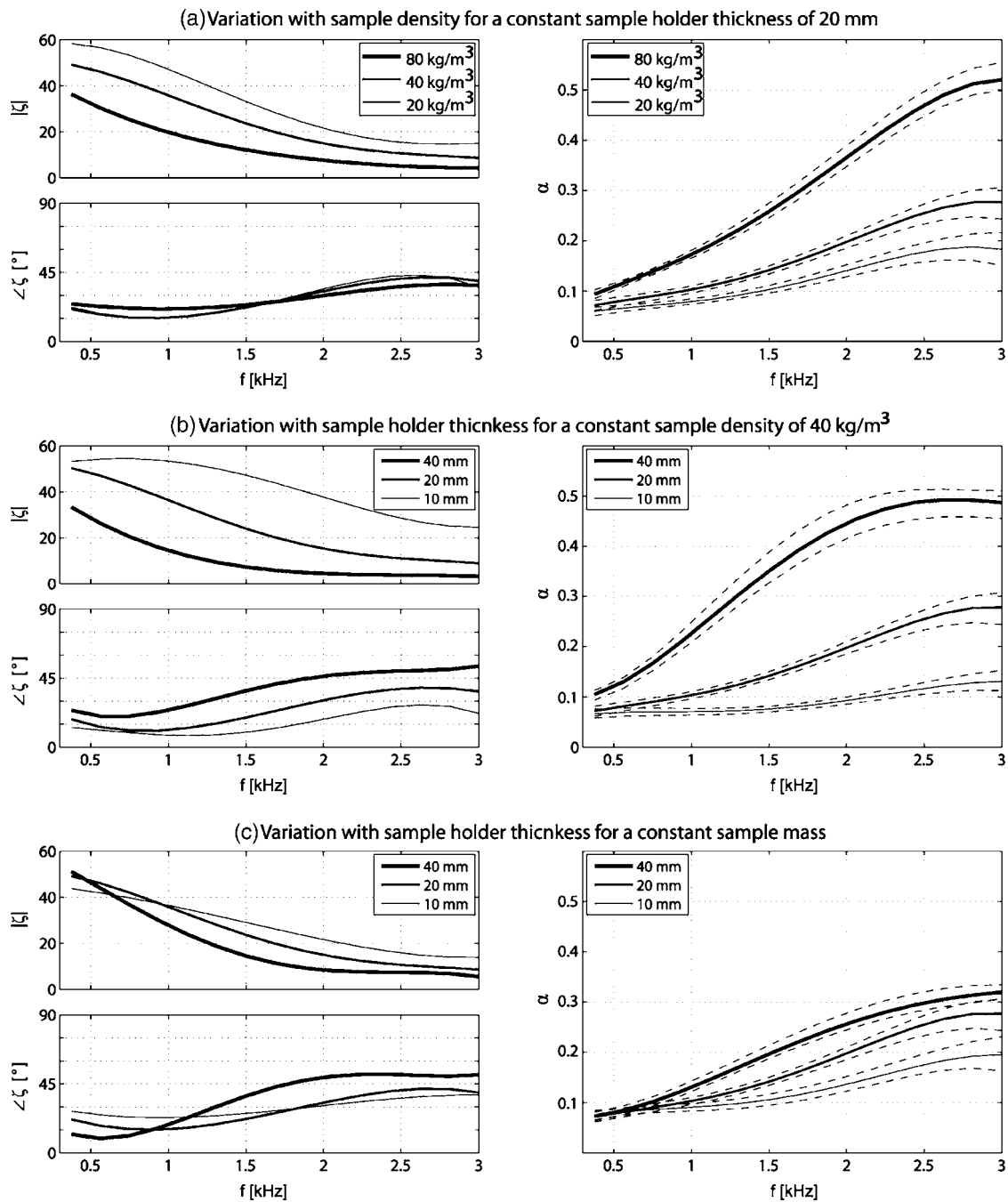
FIG. 8. Variation in equivalent impedance with (a) sample density for constant thickness, (b) sample thickness for constant density, and (c) sample thickness for constant mass (equivalent to a sample density of 40 kg/m^3 in a 20 mm sample holder).

solution for the time harmonic pressure, and the absorption coefficients provided are thus directly comparable with those from other studies. In this regard, for the limited number of tests at comparable densities, the results presented here show a good agreement with the absorption coefficients of hair provided by Katz.[8]

IV. SUMMARY AND DISCUSSION

Previous empirical studies of human auditory cues using mannequins, spheres, and boundary element methods have

shown that the addition of scalp hair introduces asymmetrical perturbations to the HRTF in the order of several dB.[3–6] These features are also noticeable in human HRTF, although little is currently known about their perceptual significance. The traditional use of sphere scattering models to account for the broad acoustic properties of the human head assume the head surface is completely rigid. Consequently, these models cannot account for, or investigate, these features. However, recent treatment of the sphere scattering problem provides an analytical model to compute the scattering of sound by a sphere with a hemispherically divided surface boundary.[7] Given the appropriate impedance data, this facilitates the in-

clusion of hair characteristics in spherical-head models, and additionally provides a method for the systematic investigation of its perceptual significance.

The current study provides the required acoustical impedance data for a pragmatic range of human hair characteristics. The legitimacy of a locally reactive surface assumption is investigated, and an appropriate boundary condition is formulated to account for the physiological composition of a human head with hair. This utilizes an equivalent impedance parameter to allow the scattering boundary to be defined at a reference plane coincident with the inner rigid surface of the head (in this case a sphere). Experimental analysis of a simple three-dimensional scattering problem illustrates that the equivalent impedance must be used in favor of the traditional surface impedance to yield physically correct pressure magnitudes. Experimental examination of the normal acoustic impedance of a representative synthetic hair material for oblique source incidence is used to show that a locally reactive boundary assumption is legitimate.

The equivalent acoustic impedance properties of a representative range of human hair samples are presented. Increasing either the bulk density or the sample thickness reduces the impedance magnitude (and increases the absorption coefficient). Increasing the sample thickness for constant mass also decreases the impedance magnitude, but to a much lesser extent. For all modifications the equivalent impedance retains a stiffness-like reactance, with the impedance phase angle in the order of 10–50°. Increasing the sample thickness produces a relative increase in the impedance phase angle. A robust correlation between the spread of results and the properties of the individual hair samples is not exhibited. Additionally, there is not a strong dependence on the relative alignment of the hair sample. The latter is confirmed by experimental results from flat-panel tests for oblique source incidence, and examination of the pressure in the proximal region around a single sphere. The absorption coefficients presented show a good general agreement with the limited other data available in the literature.

The formulation of the equivalent impedance allows the legitimate use of a locally reactive boundary for a sphere with a nonuniform boundary condition, particularly if the properties on the inner surface are required (as is the case for binaural synthesis). This formulation, however, is not consistent with typical treatments of reflection or scattering problems which generally utilize only the normal surface impedance defined at an external surface reference plane. More complex multilayered nonuniform penetrable boundaries may yield a more accurate description of the wave processes within the material layer. However, for the current problem the use of the equivalent impedance provides an adequate description of the material boundaries of interest. The comparative limits of this assumption, and the relevance of this formulation to other materials or scattering problems is outside the scope of this paper. However, a comprehensive analytic investigation into this problem may form an area of interesting further study.

[1] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," J. Acoust. Soc. Am. **104**, 3048–3058 (1998).

[2] V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, and Z. Tang, "Approximating the head-related transfer function using simple geometric models of the head and torso," J. Acoust. Soc. Am. **112**, 2053–2064 (2002).

[3] B. E. Treeby, R. M. Paurobally, and J. Pan, "The effect of impedance on interaural azimuth cues derived from a spherical head model," J. Acoust. Soc. Am. **121**, 2217–2226 (2007).

[4] B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparison to real measurements," J. Acoust. Soc. Am. **110**, 2449–2455 (2001).

[5] K. A. J. Riederer, *HRTF Analysis: Objective and Subjective Evaluation of Measured Head-Related Transfer Functions* (Ph.D. dissertation, Helsinki University of Technology, Espoo, 2005).

[6] B. E. Treeby, R. M. Paurobally, and J. Pan, "Decomposition of the HRTF from a sphere with neck and hair," in *Proceedings of the 13th International Conference on Auditory Display*, Montreal, Canada (2007).

[7] B. E. Treeby, J. Pan, and R. M. Paurobally, "Acoustic scattering by a sphere with a hemispherically split boundary condition," J. Acoust. Soc. Am. **122**, 45–56 (2007).

[8] B. F. G. Katz, "Acoustic absorption measurement of human hair and skin within the audible frequency range," J. Acoust. Soc. Am. **108**, 2238–2242 (2000).

[9] S. B. Raymond and K. Hynynen, "Acoustic transmission losses and field alterations due to human scalp hair," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 1415–1419 (2005).

[10] D. J. Sides and K. A. Mulholland, "The variation of normal layer impedance with angle of incidence," J. Sound Vib. **14**, 139–142 (1971).

[11] J. F. Allard, Y. Champoux, and J. Nicolas, "Pressure variation above a layer of absorbing material and impedance measurement at oblique incidence and low frequencies," J. Acoust. Soc. Am. **86**, 766–770 (1989).

[12] L. J. Wolfram, "Human hair: A unique physicochemical composite," J. Am. Acad. Dermatol. **48**, S106–S114 (2003).

[13] R. R. Ogle and M. J. Fox, *Atlas of Human Hair Microscopic Characteristics* (CRC, Boca Raton, FL, 1999).

[14] P. E. Hutchinson and J. R. Thompson, "The cross-sectional size and shape of human terminal scalp hair," Br. J. Dermatol. **136**, 159–165 (1997).

[15] C. R. Robbins, *Chemical and Physical Behaviour of Human Hair* (Springer-Verlag, New York, 1994).

[16] M. P. Birch, J. F. Messenger, and A. G. Messenger, "Hair density, hair diameter and the prevalence of female pattern hair loss," Br. J. Dermatol. **144**, 297–301 (2001).

[17] L. J. Goin, W. H. Mckee, and P. L. Kirk, "Human hair studies. Applications of the microdetermination of comparative density," J. Crim. Law, Criminol. Police Sci. **43**, 263–273 (1952).

[18] J. L. Dick and R. M. W. Sumner, "Development of fiber and follicle characteristics related to wool bulk in Perendale sheep over the first year of life," in *Proceedings of the New Zealand Society of Animal Production*, 314–318 (1996).

[19] R. M. W. Sumner, J. N. Clarke, T. Wuliji, N. G. Cullen, and K. G. Dodds, "Use of crossbreeding and selection in the improvement of wool bulk and fleece weight," in *6th World Congress on Genetics Applied to Livestock Production*, Armidale, Australia, 59–62 (1998).

[20] M. E. Delany and E. N. Bazley, "Acoustical properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116 (1970).

[21] N. Voronina, "Acoustic properties of fibrous materials," Appl. Acoust. **42**, 165–174 (1994).

[22] N. Voronina, "Improved empirical model of sound propagation through a fibrous material," Appl. Acoust. **48**, 121–132 (1996).

[23] K. Attenborough, "Acoustical characteristics of porous materials," Phys. Rep. **82**, 179–227 (1982).

[24] J. F. Allard and Y. Champoux, "New empirical equations for sound propa-

gation in rigid frame fibrous materials," J. Acoust. Soc. Am. **91**, 3346–3353 (1992).

[25]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw–Hill, New York, 1968).

[26]P. M. Morse and R. H. Bolt, "Sound waves in rooms," Rev. Mod. Phys. **16**, 69–150 (1944).

[27]K. U. Ingard, "Locally and nonlocally reacting flexible porous layers; a comparison of acoustical properties," J. Eng. Ind. **103**, 302–313 (1981).

[28]P. A. Martin, "Acoustic scattering by inhomogeneous spheres," J. Acoust. Soc. Am. **111**, 2013–2018 (2002).

[29]ISO 10534-2, "Determination of Sound Absorption Coefficient and Impedance in Impedance Tubes - Part 2: Transfer-Function Method," (ISO, International Organisation for Standardization, Geneva, 1998).

[30]E. A. G. Shaw, "The acoustic wave guide. I. An apparatus for the measurement of acoustic impedance using plane waves and higher order mode waves in tubes," J. Acoust. Soc. Am. **25**, 224–230 (1953).

[31]J. C. Davies and K. A. Mulholland, "An impulse method of measuring normal impedance at oblique incidence," J. Sound Vib. **67**, 135–149 (1979).

[32]J. F. Allard and B. Sieben, "The measurement of acoustic impedance at oblique incidence with two microphones," J. Sound Vib. **101**, 130–132 (1985).

[33]J.-F. Li and M. Hodgson, "Use of pseudo-random sequences and a single microphone to measure surface impedance at oblique incidence," J. Acoust. Soc. Am. **102**, 2200–2210 (1997).

[34]M. D. Dahl, E. J. Rice, and D. E. Groesbeck, "Effects of fiber motion on the acoustical behavior of an anisotropic, flexible fibrous material," J. Acoust. Soc. Am. **87**, 54–66 (1990).

[35]R. K. Cook and P. Chrzanowski, "Absorption by sound-absorbent spheres," J. Acoust. Soc. Am. **21**, 167–170 (1949).

[36]D. L. Brungart, "Acoustic scattering from a sphere with nonuniform surface impedance" (AMRL-TR-74-49, Aerospace Medical Research Laboratory, Wright Patterson Air Force Base, OH, 1974).

[37]D. A. Bies and C. H. Hansen, "Flow resistance information for acoustic design," Appl. Acoust. **13**, 357–391 (1980).

Loss-improved electroacoustical modeling of small Helmholtz resonators

Tomasz Starecki[a)]
Institute of Electronic Systems, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland

Modeling of small Helmholtz resonators based on electroacoustical analogies often results in significant disagreement with measurements, as existing models do not take into account some losses that are observed in practical implementations of such acoustical circuits, e.g., in photoacoustic Helmholtz cells. The paper presents a method which introduces loss corrections to the transmission line model, resulting in substantial improvement of simulations. Values of the loss corrections obtained from comparison of frequency responses of practically implemented resonators with computer simulations are presented in tabular and graphical form. A simple analytical function that can be used for interpolation or extrapolation of the loss corrections for other dimensions of the Helmholtz resonators is also given. Verification of such a modeling method against an open two-cavity Helmholtz structure shows very good agreement between measurements and simulations. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2773929]

I. INTRODUCTION

Helmholtz resonator is a structure that has found numerous applications. One of them is photoacoustics. Photoacoustic phenomenon relies on producing a thermal and pressure wave due to absorption of light by a substance illuminated by the light.[1] If the light intensity is modulated with an acoustic frequency, absorption of the light results in corresponding, periodic changes of pressure, which means that a sound is generated. Amplitude A_{PAS} of the sound can be described as[2]

$$A_{PAS} \propto \frac{\alpha P}{fV}, \tag{1}$$

where α is the light absorption coefficient, P is the light power, f the light intensity modulation frequency, and V the volume of the cell.

Once the photoacoustic signal is produced, its further behavior is purely acoustical. In particular, if the cell is operated at its acoustic resonance frequency, amplitude of the photoacoustic signal will be amplified by the Q factor of the cell.[2] That is why the Helmholtz structure is one of the most often used cell designs, especially in the case of photoacoustic investigation of solids.[1–3] Such a cell consists of two cavities connected with a duct. Usually the investigated sample is placed in one of the cavities and illuminated with light, while the other cavity is equipped with a sound detector (Fig. 1). One of the main advantages of the Helmholtz resonator applied as a photoacoustic cell is that its overall volume can be kept very small (a few cm^3 or even below 1 cm^3), which is quite important, as the photoacoustic signal amplitude increases against cell volume (Eq. (1)). Taking into consideration that properties of the cell directly affect sensitivity of

the whole photoacoustic setup, design of the cell should be always given special attention. A common tool used during this process is simulations of the frequency responses of the cells, so that it is possible to evaluate and compare properties of many different cells without having them manufactured, which speeds up the design process and lowers its cost. In the case of photoacoustic Helmholtz cells the most common simulation method is based on acoustoelectrical analogies.

II. ACOUSTOELECTRICAL ANALOGIES IN MODELING OF PHOTOACOUSTIC HELMHOLTZ RESONATORS

The acoustoelectrical analogies is a well-known method based on the fact that behavior of acoustical elements can be described by means of corresponding electrical components. The photoacoustic Helmholtz resonator from Fig. 1 can be then simulated in a circuit shown in Fig. 2(a), in which cavities are modeled by capacitors, and interconnecting duct by an inductor connected in series with a resistor. The first drawback of the mentioned lumped components model is that there are several sets of definitions that can be used for obtaining L and R components values, given, for instance, in the works of Morse,[4] Nolle,[5] Nordhaus and Pelzl.[6,7] But the most important is that no matter which of the mentioned definition sets is used, the model does not produce results that would be in good agreement with measured frequency responses of the resonators.[8] Although simulated resonance frequencies are usually not far away from the measured values, calculated Q factors of the resonators, which are crucial in photoacoustic applications, differ sometimes as much as about 1000 times from the measurements (if the components are defined as by Morse[4]), and even the definition set that gives the closest match often produces Q factor values 20 times greater than the measured ones.[9]

Much better results can be obtained if the interconnecting duct is represented by a transmission line (see Fig. 2(b)), which includes viscous and thermal losses.[9] Although use of

a)Electronic mail: tomi@ise.pw.edu.pl
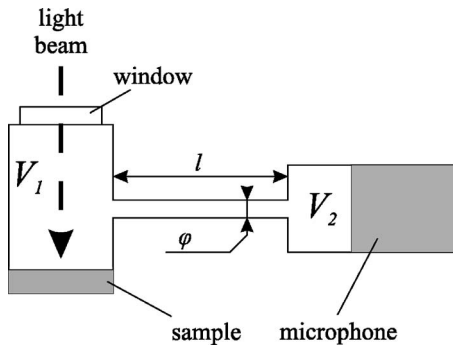
FIG. 1. Sketch of a small Helmholtz resonator applied to photoacoustic investigation of solids.

transmission line model is usually considered when length of the duct is becoming comparable to the acoustical wavelength, the model gives significant improvement also in cases of much shorter ducts, hence it should be used as a standard approach. The transmission line model is based on characteristic impedance Z_w and propagation constant Γ_w of the line which can be calculated from definitions specified, e.g., in Ref. 10. Z_w and Γ_w can be then used in conversion of the transmission line into equivalent T circuit (Fig. 2(c)).[11]

In comparison to the lumped components model, use of the transmission line model results in a much closer prediction of the resonator behavior. However, despite a significant decrease of Q factor errors, losses are still underestimated, and in some cases the ratio of simulated to measured values of Q factor still exceeds an order of magnitude (see Table I—column 4).[9]

The transmission line model can be further extended by the introduction of radiation losses R_{rad} end corrections L_{ends}, R_{ends},[5,12–15] acoustical impedance of the microphone Z_{micr},[16] viscous and thermal losses of the cavities R_{vCi} and R_{tCi},[5] etc., but the main result of these changes is an increase of complexity of the model (Fig. 3), while its properties are still far away from expectations (see Table I—column 5).[9] Such a
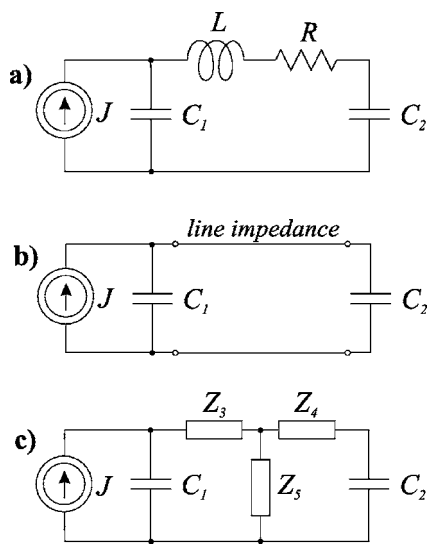
FIG. 2. Models of the resonator presented in Fig. 1, based on acoustoelectrical analogies: (a) lumped components model, (b) transmission line model, (c) transmission line model with the transmission line component replaced by an equivalent lumped T section.

difference between theoretical and measured properties of the resonators is a bit surprising, taking into consideration that many researchers have investigated properties of Helmholtz resonator and their works show usually much better agreement between theory and measurements (see, for instance, Refs. 17–20). The probable reason is that the mentioned works were conducted on the Helmholtz resonators of much larger size, while at smaller volumes other phenomena become dominant and result in a noticeable increase of the losses inside the resonators.

As a purely theoretical approach was resulting only in different versions of an unsatisfactory model, a practical workaround approach was applied. The goal was to obtain a relatively simple model that would give much closer agreement between theoretical and experimental results.

III. LOSS CORRECTIONS OF THE TRANSMISSION LINE MODEL

A fact that resonance frequencies obtained from the transmission line model were in good agreement with the measured values, and that the main problem was lack of accuracy regarding simulation of Q factors, led to the conclusion that not all the losses are included in the model. In small Helmholtz resonators viscous and thermal cavity losses are not the main reason for damping the oscillations. It is more likely that the main sources of losses are flow disturbances (due to rapid change of a cross section at the duct-cavity boundary) and interaction between the jet flowing out of the duct and being reflected by the cavity wall (or the microphone membrane) placed at a relatively small distance in front of the duct opening. Such effects can be modeled by means of resistances $R_{\ell oss}$ placed between the components that correspond to acoustical properties of the duct and the cavities (Fig. 4). The resistances $R_{\ell oss}$ can be then treated as loss corrections (similarly to the well known end corrections), which include all the losses that occur at the duct-cavity boundaries, including radiation losses, resistive components of end corrections, jet reflection losses, etc. Certainly, in order to obtain noticeable influence of the loss corrections on the frequency responses of the resonators, values of $R_{\ell oss}$ must not be much smaller than the real parts of the impedances Z_3 and Z_4. Hence, it may be convenient to express $R_{\ell oss}$ as a function of the impedances Z_3 and Z_4. In order to simplify the model it was assumed that relationship between $R_{\ell oss}$ and Z_3, Z_4 for a given resonator is fixed, so that

$$R_{\ell oss} = \gamma \, \mathrm{Re}(Z_3) = \gamma \, \mathrm{Re}(Z_4). \tag{2}$$

Such an approach is a modified concept of correction factors presented briefly in Ref. 21, where one can also find a detailed description of how to calculate values of the components used in the models presented in Figs. 2 and 3 and of a routine used for extraction of the correction factors. In order to evaluate coefficients γ (retrieved similarly to the mentioned correction factors), frequency responses of 48 different combinations of resonator dimensions were measured and compared to simulated responses. The resonators had 2.0 cm^3 sample cavity, the microphone cavity was 0.5, 1.0, 1.5, and 2.0 cm^3, and the duct between the cavities was 2.0,

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Tomasz Starecki: Modeling of small Helmholtz resonators   2119

TABLE I. Values of the coefficients γ vs resonator dimensions extracted from the measurements of 48 different Helmholtz resonators and theoretical to measured ratio of Q factors of these resonators (theoretical Q factor values calculated from the models given in Figs. 2(c) and Fig. 3).

Duct diameter [mm]	Duct length [cm]	Microphone cavity volume [cm^3]	Q_{theor}/Q_{meas} (Fig. 2(c))	Q_{theor}/Q_{meas} (Fig. 3)	γ
1	2	0.5	2.00	1.95	0.30
1	2	1	1.49	1.48	0.23
1	2	1.5	1.32	1.31	0.16
1	2	2	1.17	1.16	0.15
1	3	0.5	2.03	1.96	0.22
1	3	1	1.50	1.48	0.15
1	3	1.5	1.39	1.38	0.25
1	3	2	1.23	1.22	0.10
1	4	0.5	2.14	2.07	0.25
1	4	1	1.69	1.66	0.13
1	4	1.5	1.42	1.40	0.15
1	4	2	1.10	1.09	0.10
2	2	0.5	4.08	4.02	0.74
2	2	1	2.33	2.31	0.37
2	2	1.5	1.94	1.92	0.30
2	2	2	1.66	1.65	0.22
2	3	0.5	3.36	3.27	0.61
2.	3	1	2.42	2.39	0.35
2	3	1.5	2.00	1.99	0.36
2	3	2	1.69	1.68	0.21
2	4	0.5	3.30	3.23	0.58
2	4	1	2.40	2.37	0.33
2	4	1.5	2.03	2.01	0.28
2	4	2	1.70	1.69	0.20
3	2	0.5	7.56	7.42	1.40
3	2	1	3.67	3.64	0.63
3	2	1.5	2.81	2.79	0.46
3	2	2	2.18	2.17	0.33
3	3	0.5	5.18	5.06	1.13
3	3	1	3.48	3.45	0.59
3	3	1.5	2.70	2.69	0.48
3	3	2	2.13	2.13	0.34
3	4	0.5	5.04	4.97	1.03
3	4	1	3.51	3.47	0.55
3	4	1.5	2.64	2.63	0.42
3	4	2	2.16	2.15	0.34
4	2	0.5	13.17	12.91	2.00
4	2	1	5.67	5.62	0.94
4	2	1.5	3.98	3.96	0.62
4	2	2	2.92	2.91	0.42
4	3	0.5	8.12	7.96	1.70
4	3	1	4.96	4.91	0.92
4	3	1.5	3.66	3.63	0.69
4	3	2	2.84	2.83	0.47
4	4	0.5	7.24	7.11	1.49
4	4	1	4.97	4.92	0.84
4	4	1.5	3.62	3.59	0.63
4	4	2	2.74	2.73	0.49

3.0 or 4.0 cm long and 1.0, 2.0, 3.0 or 4.0 mm in diameter. The obtained values of coefficients γ are presented in Table I and in Fig. 5.

When analyzing the above set of the coefficients γ, it can be easily noticed that they increase with the interconnecting duct diameter and against microphone cavity vol-
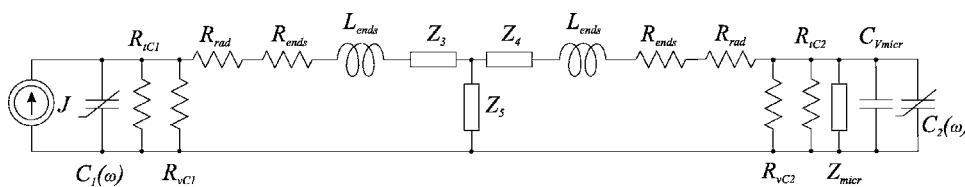
FIG. 3. Enhanced transmission line model including additional loss mechanisms, e.g., radiation impedance, end corrections, microphone impedance, etc.

Tomasz Starecki: Modeling of small Helmholtz resonators
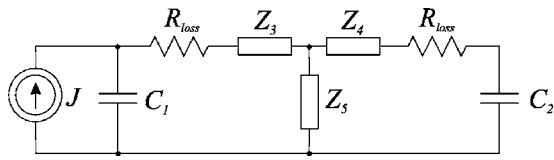
FIG. 4. Simple transmission line model with loss corrections ($R_{\ell\text{oss}}$).

ume, while they are not much influenced by the duct length. This confirms the previous assumption that the transmission line model quite precisely describes losses inside the duct, and that the main loss sources which were not included in the previous models are located outside the duct.

Although a set of coefficients γ is already of some help in the process of designing, having an analytical function describing such coefficients would be much more convenient, especially if applied for modeling purposes. Hence, some further work concentrated on development of such an analytical description. Taking into consideration, that values from Table I show a slightly nonlinear relationship of γ versus duct diameter φ, parabolic approximation was used, so that

$$\gamma = A\varphi + B\varphi^2. \tag{3}$$

In order to find description for A and B the following routine was applied. At first all 12 point sets from Fig. 5 were approximated by individual parabolic functions (Fig. 6(a)). The functions were then analyzed in order to obtain a common form as in the Eq. (3) with some efforts made towards compromise between a relatively simple definition of A and B and possibly close fit to the given experimental data from Table I, which finally resulted in

$$A = 0.0001 \cdot (1 - 8\,\ell\,)\frac{V_1 + V_2}{V_1 V_2}, \tag{4}$$
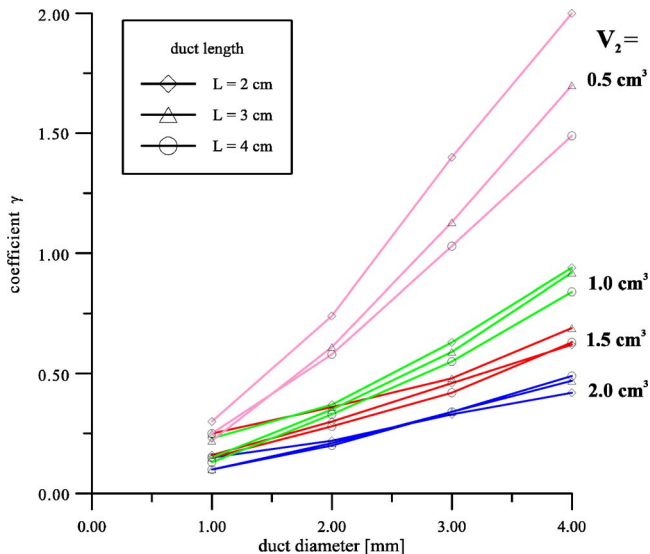
FIG. 5. (Color online) Coefficients γ obtained from measurements of 48 different photoacoustic Helmholtz resonators.
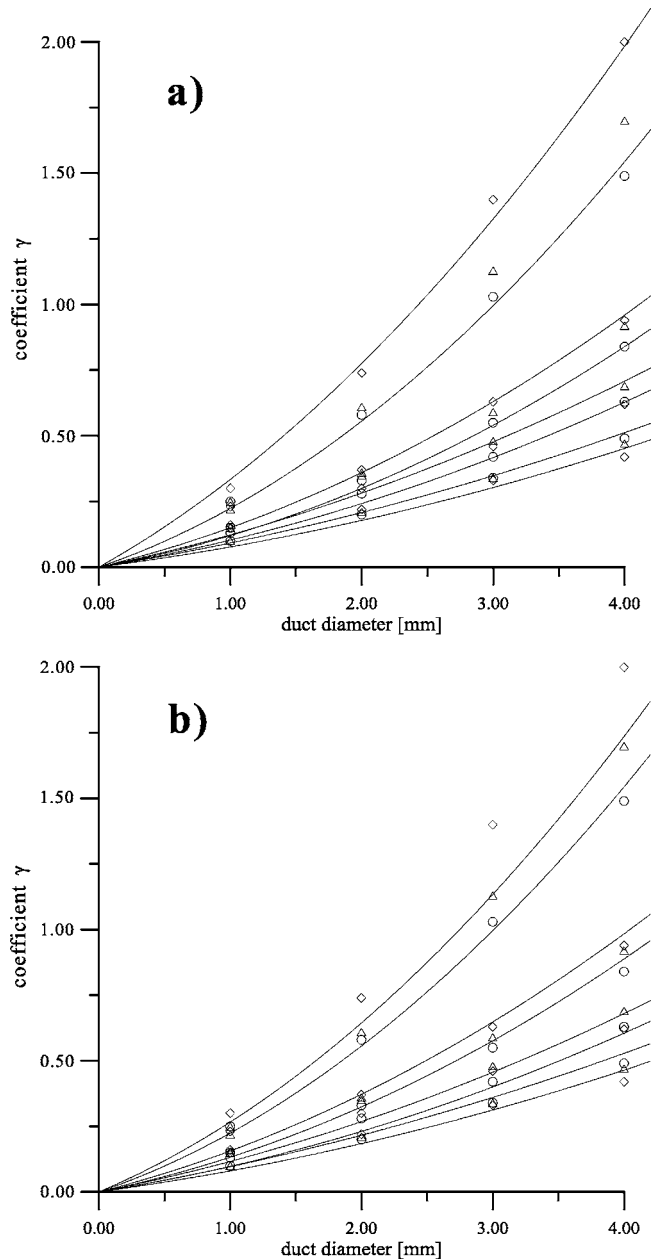
FIG. 6. Approximation of coefficients γ presented in Fig. 5 by analytical functions: (a) with every point set approximated by an individual function, (b) with a single function used for approximation of all the points.

$$B = \sum_{i=1}^{2} 96{,}000 \cdot 2^{-2{,}000{,}000 V_i}, \tag{5}$$

where Eq. (4) is valid only for duct shorter than 12.5 cm (for greater ℓ values, A should be substituted by zero). In the above equations (Eqs. (2)–(5)) standard SI units should be used (the length and diameter of the duct ℓ should be given in meters, volumes V_1, and V_2 at both ends of the duct in m^3). In comparison to individual parabolic functions from Fig. 6(a) the above definitions give a little bit worse fit to the experimental data (Fig. 6(b)), but the main character of the relationship is preserved.

The presented loss-improved transmission line components were verified against an open photoacoustic Helmholtz cell. Structure of the cell is shown in Fig. 7(a). In order to

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Tomasz Starecki: Modeling of small Helmholtz resonators   2121
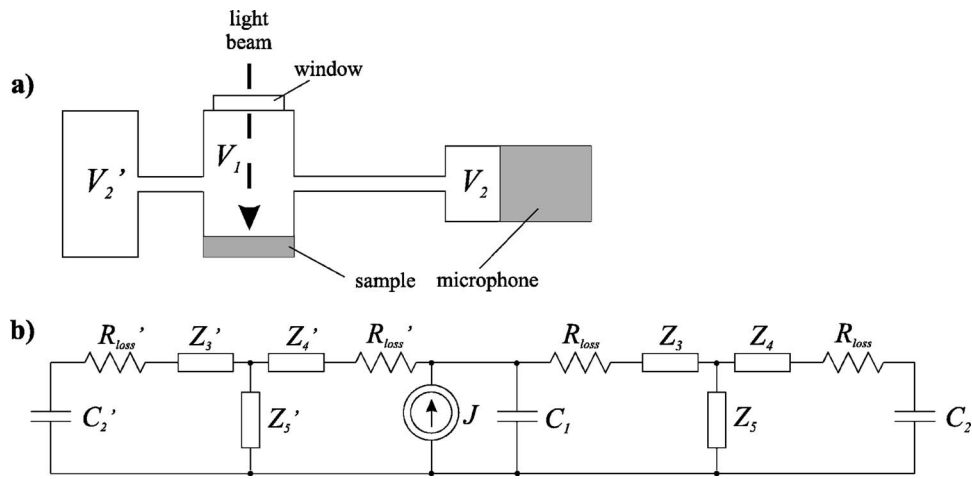
FIG. 7. Triple-cavity Helmholtz resonator (a) and its transmission line model with loss corrections (b) used for simulation of an open dual-cavity Helmholtz resonator (for this purpose volume V_2' was assigned value of several m³).

use a similar approach as in dual cavity closed resonator, the external space was treated as an additional cavity V_2' of a large volume (e.g., a few m³). Such a resonator can be modeled in a circuit given in Fig. 7(b). In the measured resonator the sample cavity was 2.0 cm³ in volume, microphone cavity was 1.5 cm³, duct between the two was 3.0 cm in length and 3.0 mm in diameter, while the connection between the sample cavity and external space had a form of two parallel ducts: 5.0 cm in length and 2.0 mm in diameter each. Figure 8 presents measured frequency response of the resonator (black dots) compared with results calculated from the transmission line model without corrections (solid line) and with loss corrections (based on Eqs. (2)–(5); dashed line). The frequency response has two resonance peaks. The one at the higher frequencies results mainly from interaction of the two smaller cavities and the interconnecting duct (that were of
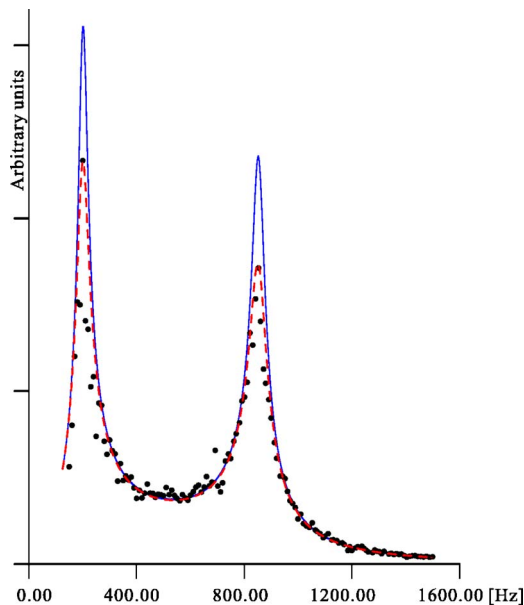
dimensions identical to one of the resonators measured and used for the purpose of evaluation of the gamma coefficients). The other resonance peak results mainly from the interaction of the third (large) cavity that was connected with the sample cavity with the ducts that were longer than these used for the purpose of evaluation of the gamma coefficients. Good agreement between the simulations and measurements proves that the analytical function describing the gamma coefficients was implemented correctly, because the model works well in the case when partial geometry of the resonator is similar to one of these from which the gamma coefficients were derived (the right resonance peak). It shows that the model also gives very good results in the case when it was used with extrapolation of the loss corrections, when the geometry was different from the previously measured structures (the left resonance peak). It is clearly visible, that although both models (without and with loss corrections) properly reflect actual resonance frequencies, amplitudes measured at the frequencies close to resonances are in good agreement only with the model including loss corrections. Fitting quality of both models was compared by calculation of mean absolute percentage error and root averaged squared error. For the experimental data as given in Fig. 8 the mentioned errors were at the level of 10% and 14% for the best fit of the loss-improved model, while for the best fit of the model without loss corrections the errors were significantly higher (respectively 19% and over 60%), which proves much higher fitting quality of the loss-improved model.

Finally, it should be noticed, that if none of the cavities is small, the values of A and B are negligible and the proposed loss-improved model approaches the standard transmission line model from Fig. 2(c). This means that the presented model can be treated as an enhancement which preserves features of the standard transmission line model for the Helmholtz resonators of standard (not very small) size, but which substantially improves modeling results for the resonators of reduced size. The proposed loss-correction model was intended for modeling of Helmholtz resonators

FIG. 8. (Color online) Comparison of measurements of the open two-cavity Helmholtz resonator (black dots) with simulations based on standard transmission line model (solid line) and transmission line model with loss corrections (dashed line).

(including multi-cavity structures) with cavities of at least 0.5 cm^3 and ducts with diameter of $1-4$ mm, not shorter than 2 cm.

IV. CONCLUSIONS

Some cited works and presented comparison of the measured and simulated frequency responses of a small Helmholtz resonator showed that the standard transmission line model (even with some enhancements like radiation impedance, microphone impedance, end corrections, etc.) based on acoustoelectrical analogies does not produce satisfactory results due to underestimation of losses that occur in the resonator. Proposed improvement of the model was based on the concept of additional loss correction components placed at the ends of the transmission line. Loss corrections were obtained experimentally from a reasonable number of two-cavity Helmholtz resonators and approached by an analytical function, that can be used for interpolation and extrapolation of the correction factors for resonators of some other dimensions. Verification of the presented model with loss corrections against an open multi-cavity Helmholtz resonator showed very good agreement between measured and simulated frequency responses, which leads to the conclusion that the presented approach has a good chance of producing promising results not only in the case of simple two-cavity resonators, but also in the case of complex, multi-cavity small acoustic structures.

[1] Y.-H. Pao, *Optoacoustic Spectroscopy and Detection* (Academic, New York), Chap. 8.

[2] A. Miklós, P. Hess, and Z. Bozóki, "Application of acoustic resonators in photoacoustic trace gas analysis and metrology," Rev. Sci. Instrum. **72**, 1937–1955 (2001).

[3] V. P. Zahrov and V. S. Letokhov, *Laser Optoacoustic Spectroscopy*, Springer Series in Optical Sciences (Springer, Berlin, 1986), Vol. **37**, Chap. 5.3.

[4] P. M. Morse, *Vibration and Sound* (McGraw–Hill, New York, 1948), pp. 234–235.

[5] A. W. Nolle, "Small-signal impedance of short tubes," J. Acoust. Soc. Am. **25**, 32–39 (1953).

[6] O. Nordhaus and J. Pelzl, "Frequency dependence of resonant photoacoustic cells: The extended Helmholtz resonator," Appl. Phys. **25**, 221–229 (1981).

[7] J. Pelzl, K. Klein, and O. Nordhaus, "Extended Helmholtz resonator in low-temperature photoacoustic spectroscopy," Appl. Opt. **21**, 94–99 (1982).

[8] T. Starecki, "Modeling of photoacoustic Helmholtz by means of acoustic-electrical analogies," Electronics Telecommunications Quaterly **39**, 307–312 (1993).

[9] T. Starecki, *Analiza Porównawcza Modeli Komory Helmholtza do Przyrządów Pomiarowych Typu PAS* (*Comparitive Analysis of Photoacoustic Helmholtz Cell Models*-in Polish) (Ph.D. dissertion, Warsaw University of Technology, Warsaw (1994).

[10] A. H. Benade, "On the propagation of sound waves in a cylindrical conduit," J. Acoust. Soc. Am. **44**, 616–623 (1968).

[11] F. B. Daniels, "On the propagation of sound waves in a cylindrical conduit," J. Acoust. Soc. Am. **22**, 563–564 (1950).

[12] U. Ingard, "On the theory and design of acoustic resonators," J. Acoust. Soc. Am. **25**, 1037–1061 (1953).

[13] R. W. Troke, "Tube-cavity resonance," J. Acoust. Soc. Am. **44**, 684–688 (1968).

[14] J. B. Mehl, "Greenspan acoustic viscometer: Numerical calculations of fields and duct-end effects," J. Acoust. Soc. Am. **106**, 73–82 (1999).

[15] K. A. Gillis, J. B. Mehl, and M. R. Moldover, "Theory of Greenspan viscometer," J. Acoust. Soc. Am. **114**, 166–173 (2003).

[16] A. J. Zuckerwar, "Theoretical responses of condenser microphones." J. Acoust. Soc. Am. **64**, 1278–1285 (1978).

[17] A. Selamet, P. M. Radavich, N. S. Dickey, and J. M. Novak, "Circular concentric Helmholtz resonators," J. Acoust. Soc. Am. **101**, 41–51 (1997).

[18] M. Moloney, "Quality factors and conductances in Helmholtz resonators," Am. J. Phys. **72**, 1035–1039 (2004).

[19] T. A. Johansson and M. Kleiner, "Theory and experiments on the coupling of two Helmholtz resonators," J. Acoust. Soc. Am. **110**, 1315–1328 (2001).

[20] J. Wu and I. Rudnick, "Measurements of the nonlinear tuning curves of Helmhlotz resonators," J. Acoust. Soc. Am. **80**, 1419–1422 (1986).

[21] T. Starecki, "Practical improvements of modeling of photoacoustic Helmholtz cells," Proc. SPIE **6159**, 653–658 (2006).

Further studies on the dual-resonance nonlinear filter model of cochlear frequency selectivity: Responses to tones[a)]

Alberto Lopez-Najera[b)]
Facultad de Medicina, Universidad de Castilla-La Mancha, C/ Almansa, No. 14, 02006 Albacete, Spain

Enrique A. Lopez-Poveda
*Instituto de Neurociencias de Castilla y León, Universidad de Salamanca, Avda. Alfonso X "El Sabio"
s/n, 37007 Salamanca, Spain*

Ray Meddis
Department of Psychology, University of Essex, Colchester CO4 3SQ, United Kingdom

A number of phenomenological models that simulate the response of the basilar membrane motion can reproduce a range of complex features observed in animal measurements over different sites along its cochlea. The present report shows a detailed analysis of the responses to tones of an improved model based on a dual-resonance nonlinear filter. The improvement consists in adding a third path formed by a linear gain and an all-pass filter. This improvement allows the model to reproduce the gain and phase plateaus observed empirically at frequencies above the best frequency. The middle ear was simulated by using a digital filter based on the empirical impulse response of the chinchilla stapes. The improved algorithm is evaluated against observations of basilar membrane responses to tones at seven different sites along the chinchilla cochlear partition. This is the first time that a whole set of animal observations using the same technique has been available in one species for modeling. The resulting model was able to simulate amplitude and phase responses to tones from basal to apical sites. Linear regression across the optimized parameters for seven different sites was used to generate a complete filterbank. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2769627]

I. INTRODUCTION

A number of phenomenological models that simulate basilar membrane (BM) responses can reproduce a range of complex features observed in animal measurements over different sites along the cochlea (Giguere and Woodland, 1994; Goldstein, 1990; 1995; Irino and Patterson, 2001; Meddis *et al.*, 2001; Robert and Eriksson, 1999; Zhang *et al.*, 2001; Zilany and Bruce, 2006; Lopez-Poveda, 2005). For example, the dual resonance nonlinear (DRNL) filter of Meddis *et al.* has been shown to be able to simulate a wide range of phenomena characteristic of the vibration of the cochlear partition. These include nonlinear (compressed) responses, changes in effective bandwidth (BW) with signal level, the response to click-stimuli at different levels, two-tone suppression, and the generation of distortion products.

Unfortunately, this model is unable to simulate BM responses to frequencies well above the characteristic frequency (CF), the so-called "plateau response," as measured by Ruggero (1997). The deviation is clearly seen by comparing Figs. 1C and 1D in Meddis *et al.* (2001); the model substantially underestimates the experimental response for 17-kHz stimuli at a site whose CF is 10 kHz. An amendment to the DRNL filter is proposed in the following that addresses this issue and improves the fit at frequencies above CF. These responses may be relatively unimportant at low levels but they substantially change the nature of the mechanical filtering at high signal levels when the filters become extremely wide.

Phenomenological models of the mechanical response have also faced the lack of systematic observations for a wide range of CFs. Difficulties associated with making these measurements have restricted the number of sites along the partition that can be studied. Until recently, model parameters could only be adjusted based on experimental data for three cochlear sites with BFs of 800, 8000, and 18 000 Hz across different species. Any attempt to derive a filterbank representing the whole cochlea necessarily involved a considerable degree of interpolation between these points or a dependence on auditory nerve data to fill the gaps (Sumner *et al.*, 2003). Over time, however, more observations have been made and modelers now have access to experimental BM responses for seven cochlear regions with CFs of approximately 0.8, 5.5, 7.25, 9.75, 10, 12, and 14 kHz (Rhode and Cooper, 1996; Rhode and Recio, 2000; Ruggero *et al.*, 1997) in a single species, the chinchilla. While still far from complete, the new data provide the opportunity to reassess the

---

A

B

model parameters. More phase data are also now available and this also permits a reexamination of the phase response of the DRNL filter.

When evaluating models of the response of the cochlear partition, it is important to take into account the frequency response of the outer-middle ear (Cheatham and Dallos, 2001). Fortunately, the experimental impulse response (IR) of the stapes is now available (Rhode and Recio, 2000, private communication). This can be represented as a Finite Impulse Response (FIR) filter so that middle-ear response can be computer-simulated with some confidence. This filter will also be presented and evaluated in the following.

The main improvement to the DRNL filter of Meddis *et al.* (2001) is the introduction of a third signal processing path, making the filter a triple-path nonlinear (TRNL) model. The third path is a linear, low-gain, all-pass filter. Its gain is low enough so that it is only effective at high signal levels and its frequency response flat to act as an all-pass filter (Robles and Ruggero, 2001). The third filter is addressing the "plateau response" observed at frequencies well above CF (Ruggero *et al.*, 1997). When combined with its low gain, this filter is effective only outside the bandpass of the existing nonlinear and linear bandpass filters. The parameters of the TRNL filter have been adjusted to account for the bulk of BM responses to pure tones that are now available. The most important adjustment, with respect to the original DRNL filter, is an increase in the number of gammatone and low-pass filters required to fit the new data particularly for high frequencies.

II. MODEL DESCRIPTION

The general scheme of the model implementation is illustrated in Fig. 1. A pressure wave form (in units of pascal) is the input to the algorithm. The middle-ear (ME) filter pro-

duces stapes velocity (in m/s), which is the input to the TRNL filter. To preserve as many aspects of the experimental stapes response as possible, the ME filter was implemented as a 256-point FIR filter whose coefficients (Oppenheim and Schafer, 1999) were equal[1] to the empirical sensitivity IR of the chinchilla stapes shown in Fig. 2(a) (case CB058 in Rhode and Recio, 2000, private communication). The gain and phase responses of this filter are shown in Fig. 2(b).

The TRNL filter was implemented and evaluated digitally in the time domain. A new parallel path was added to the original DRNL (Meddis *et al.*, 2001) consisting of a linear, zero-phase, all-pass filter. High frequency amplitude and phase plateaus observed in BM tonal responses (Robles and Ruggero, 2001) were not reproduced by the original DRNL filter. The idea by Robles and Ruggero (2001, p. 1313) that the plateaus "[…] reflect, more or less directly […] stapes motion […]" suggested the implementation of the new third path as a zero-phase, all-pass filter. Phase responses measured with a sensor placed at 300 μm from the BM in the scala vestibuli (Olson, 1998) also support this idea.

Fundamentals for implementing the original DRNL filter can be found in the Appendix of Lopez-Poveda and Meddis, (2001). The zero-phase, all-pass filter in the new third path was implemented digitally by filtering the input signal in both the forward and reverse directions through a second-order Butterworth low-pass filter with a very high cut-off frequency, almost equal to the Nyquist frequency (Oppenheim and Schafer, 1999). This was achieved using MATLAB's filtfilt function. The output signal from the TRNL filter was the sample-by-sample summation of the output signals from the DRNL filter and the new third path. The gain (scalar) of the third path, k, was free to vary above zero, but was always lower than the gain of the linear path, g, of the DRNL filter. For this reason, the contribution of the third path is promi-
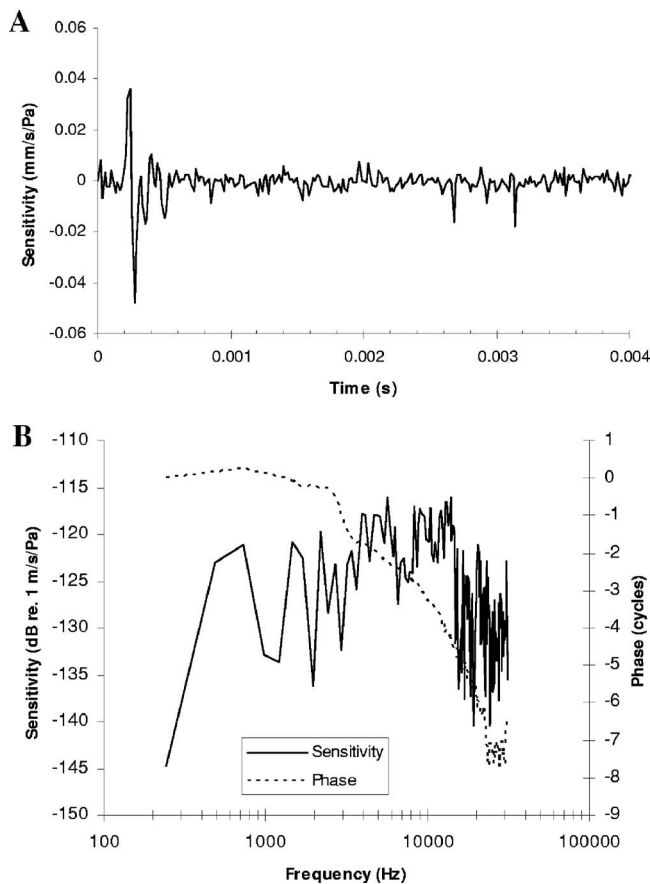
FIG. 2. Transfer characteristics of the ME filter. (A) Experimental impulse of preparation CB058 (Rhode, private communication). This was used as the coefficients for a FIR filter that simulates the ME stage. (B) Sensitivity and phase response of the FIR filter used.

nent only at high levels and for frequencies above the center frequencies of the filters in the linear and the nonlinear path, where the output of the linear and the nonlinear paths are attenuated.

Another difference between the TRNL filter and the DRNL filter is the number of gammatones (GT) and low-pass (LP) filters in cascade. The number of GTs in the linear path was increased to 5, and the number of LP filters was fixed equal to 7. The number of LP filters in the nonlinear path was also increased to 4. A greater number of filters makes the attenuation above CF steeper, where the contribution from the third path is more relevant to the total TRNL-filter output.

A. Model parameters

The parameters of the filters, shown in Table I, were optimized by manually fitting the model response to the experimental BM velocity input/output (I/O) functions as in the original DRNL paper (Meddis et al., 2001), but also taking into account the experimental frequency, phase, and IRs (not shown).

The procedure for obtaining the parameters was as described in Meddis et al. (2001), that is, we assumed that the output from the nonlinear path dominates the output of the TRNL filter at very low stimulus level near its CF, and the linear-path output dominates the TRNL output at high levels.

These two conditions allow obtaining the parameters for the nonlinear and linear paths, respectively. The gain, k, of the new third path was obtained by fitting the output of the TRNL filter to the frequency and phase plateaus observed in the experimental responses for frequencies above CF, for which the output from the linear and the nonlinear paths is negligible.

The number of cascaded filters in each path was obtained by fitting the model phase-shift responses at low and high stimulus levels to the data (Ruggero et al., 1997; Rhode and Recio, 2001).

B. Evaluation of the model

The model was implemented in MATLAB. Digital sinusoidal tones (sampling frequency of 62 500 Hz) were used as the input, simulating exactly the same stimuli used in the experiments. Input signals had a total duration of 30 ms, including 2-ms rise/fall raised-cosine ramps. Signal levels from 0 to 100 dB SPL were considered. The phase of the model responses was calculated using a sine-wave fitting algorithm (Händel, 2000). The amplitude of model responses were calculated as the peak velocity (m/s) during the last half of the signal. All filters were implemented using the functions available in MATLAB and in its Signal Processing Toolbox.

III. RESULTS

Parameters were optimized for those cochlear sites for which experimental data were available for chinchilla (see Table I). In what follows, model results are compared directly with experimental responses. Only those results and conditions for which experimental observations exist were reproduced with the model: I/O and frequency response functions were reproduced for all CFs; phase versus frequency responses were modeled only for CFs of 0.8, 7.25, and 12 kHz; and level-dependent phase responses were modeled for CFs of 7.25 and 10 kHz.

A. Frequency response and input/output curves

Figures 3(a), 3(c), and 3(e) illustrate experimental (left panels) and model (right panels) frequency responses for CFs of 0.8, 5.5, and 7.25 kHz. Figures 3(b), 3(d), and 3(f) show the same data plotted as I/O curves. Similarly, frequency responses for CFs of 9.75, 10, 12, and 14 kHz are plotted in Figs. 4(a), 4(c), 4(e), and 4(g) and the corresponding I/O functions in Figs. 4(b), 4(d), 4(f), and 4(h). By plotting frequency response and I/O functions, the reader can easily visualize filter shapes, linearity, and compression.

1. 800-Hz site

Figures 3(a) and 3(b) show sensitivity and I/O curves, respectively, for a cochlear region with a CF $\sim$ 800 Hz [Rhode and Cooper (1996), case CH16]. No ME filter was used to model these data because there were no detailed data available to simulate its response around 800 Hz. Nonlinear responses can be seen only for frequencies of 800, 900, and 1000 Hz around 50 dB SPL as a bend in the I/O functions.

TABLE I. TRNL algorithm parameters used throughout for reproducing the animal observations considered in this report. The top three rows inform of the cochlear site (CF), the preparation number, and the study from which experimental data were taken. Studies: 1: Rhode and Cooper (1996); 2: Rhode and Recio (2000); 3: Ruggero et al. (1997).

	CF	800	5500	7250	9750	10 000	12 000	14 000
	case	CH16	CB58	CB61	CB24	L113	CB04	CB21
	study	1	2	2	2	3	2	2
Linear								
GT cascade		5	5	5	5	5	5	5
LP cascade		7	7	7	7	7	7	7
CF_{lin}		750	5000	7 400	9000	9 000	11 000	13 000
BW_{lin}		450	3000	2 500	3000	3 500	5 000	4 000
LP_{lin}		750	6000	7 400	9000	8 800	12 000	13 500
Gain, g		500	190	3 000	300	500	500	350
Nonlinear								
GT cascade		3	3	3	3	3	3	3
LP cascade		4	4	4	4	4	4	4
CF_{nl}		730	5850	7 800	9800	10 000	12 000	15 000
BW_{nl}		350	1800	2 275	1650	1 800	2 000	3 200
LP_{nl}		730	5850	7800	9800	10 000	12 000	15 000
Gain, a		850	3000	15 000	9000	15 000	22 500	3 000
Gain, b		0.03	0.04	0.06	0.05	0.06	0.07	0.045
Exponent, c		0.25	0.25	0.25	0.25	0.25	0.25	0.25
Linear all-pass								
Gain, k		10	0.4	20	1	2	20	20

This effect is reproduced both quantitatively and qualitatively with the model. The model parameters (see Table I, case CH16) were chosen so that the center frequency of the GT filters in the linear path were slightly higher that those of the GT filters in the nonlinear path. The gains for the two paths were also similar. The BWs of the GT filters in the linear and the nonlinear paths were similar. However, the linear-path GT filters were slightly wider than those of the nonlinear path. Compressive responses are visible only around the center frequency of the GT filters in the nonlinear path and between 50 and 70 dB SPL, where the output from the linear path dominates the total filter output. The frequency response of the model with the same model parameters but *without* the new third path at 90 dB SPL is also shown (thin line, labeled as "No 3rd"). In this case, a plateau does not occur either in the model or in the experimental responses at high frequencies. The value of the gain of the third path (parameter k in Table I) is therefore ambiguous.

2. 5.5-kHz *site*

Figures 3(c) and 3(d) compare model results and experimental data (Rhode and Recio, 2000, case CB58). In this case, the ME filter used in the model was based on the stapes IR of the same animal. Nonlinear responses occur for all frequencies shown and for levels between 30 and 80 dB SPL, and the model is able to reproduce them. Experimental and model frequency responses are comparable and the plateau is produced only when the third path is used. CF_{lin} and CF_{nl} are different, with $CF_{nl} > CF_{lin}$. The BWs are also different; $BW_{lin}=3$ kHz and $BW_{nl}=1.8$ kHz in order to reproduce the level-dependent frequency responses and the shifts of BF between low and high input levels.

3. 7.25-kHz *site*

Figures 3(e) and 3(f) illustrate experimental (Rhode and Recio, 2000, case CB61) and model responses for a cochlear site with a CF=7.25 kHz. When the new third path is omitted (line labeled "No 3rd") the frequency response does not reach a plateau at frequencies well above 8 kHz. The center frequency of the linear-path GT filters (7.4 kHz) is lower than that of the nonlinear-path filters (7.8 kHz), but the BW of the linear path (2.5 kHz) is larger than that of the nonlinear path (2.275 kHz). These parameters make it possible that the model level-dependent frequency response matches that of the data. Nonlinear responses occur in the model for frequencies between 6.25 and 10 kHz, as occurs in experimental data.

4. 9.75-kHz *site*

Figure 4(a) compares experimental (Rhode and Recio, 2000, case CB24) and model frequency responses. The ME filter used produces a minimum close to 8.5 kHz (line labeled "stapes") and thus affects the model output by producing a nondesirable minimum around that frequency. The model does not reproduce correctly the shape of the experimental filter from low to high levels. This occurs because the center frequency of the linear- and nonlinear-path GT filters are similar (9 and 9.8 kHz, respectively). Corresponding I/O functions are plotted in Fig. 4(b). Nonlinear responses can be seen for all frequencies. Note that the model underestimates the BM velocity at 8 kHz, but this is due to the minimum observed in the ME-filter sensitivity. Because of this, in this case, the fit to experimental data is qualitative rather than quantitative.
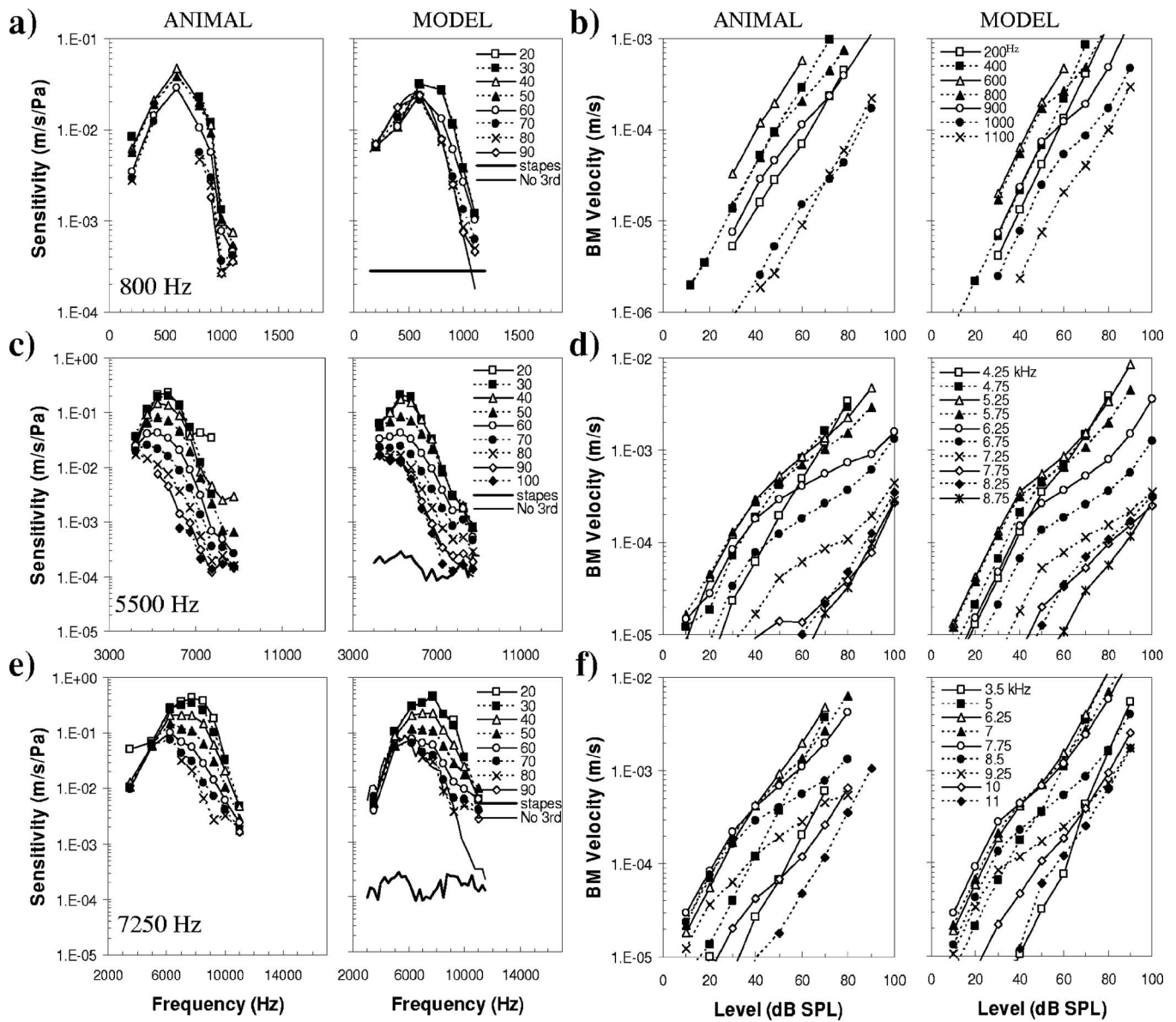
FIG. 3. Comparison of model and animal results for cochlear sites with CFs of 800 [data from Rhode and Cooper (1996), animal CH16], 5500, and 7250 Hz [data from Rhode and Recio (2000) animals CB58 and CB61 respectively]. (a), (c), and (e) Experimental and modeled frequency responses for different stimulus levels (as indicated by the insets in dB SPL). Also shown are the frequency response of the model stapes (thick line) and the response of the model without the third path at the highest level (thin line). Left panels: animal observations. Right panels: Model results corresponding to animal observations immediately to the left. (b), (d), and (f) Experimental and model input/output curves. Left panels: animal observations. Right panels: Model results corresponding to animal observations immediately to the left. The insets indicate stimulus frequency in hertz.

5. 10-kHz *site*

The experimental data for this site is perhaps the most complete set of BM measurements available in literature (Ruggero *et al.*, 1997; Recio *et al.*, 1998). Further details and analysis on modeling these experimental responses to tones and clicks using different ME-filter implementations can be found in Lopez-Najera *et al.* (2005). The model produces filter shapes similar to those observed experimentally [Fig. 4(c)]. Some of the differences possibly reflect the fact that the frequency response of the ME filter used in the model differs from that of the animal from which BM responses were measured [compare the thick continuous lines in the two panels of Fig. 4(c)]. Figure 4(d) compares corresponding I/O functions showing nonlinear responses for frequencies between 9 and 12 kHz in both animal and model responses.

Animal I/O curves show a notch at 10 kHz and 100 dB SPL that is not produced by the model as it was by Meddis *et al.* (2001). The model produces notches when the outputs from the linear and the nonlinear paths have similar amplitudes but different phases. This condition is not met at 10 kHz with the parameters shown in Table I (case L113). Note that the plateau in the frequency response is only reproduced by the model when the third path is considered.

6. 12-kHz *site*

Figures 4(e) and 4(f) compare experimental and model sensitivity and I/O functions for case CB04 of Rhode and Recio (2000). The model simulates the experimental data
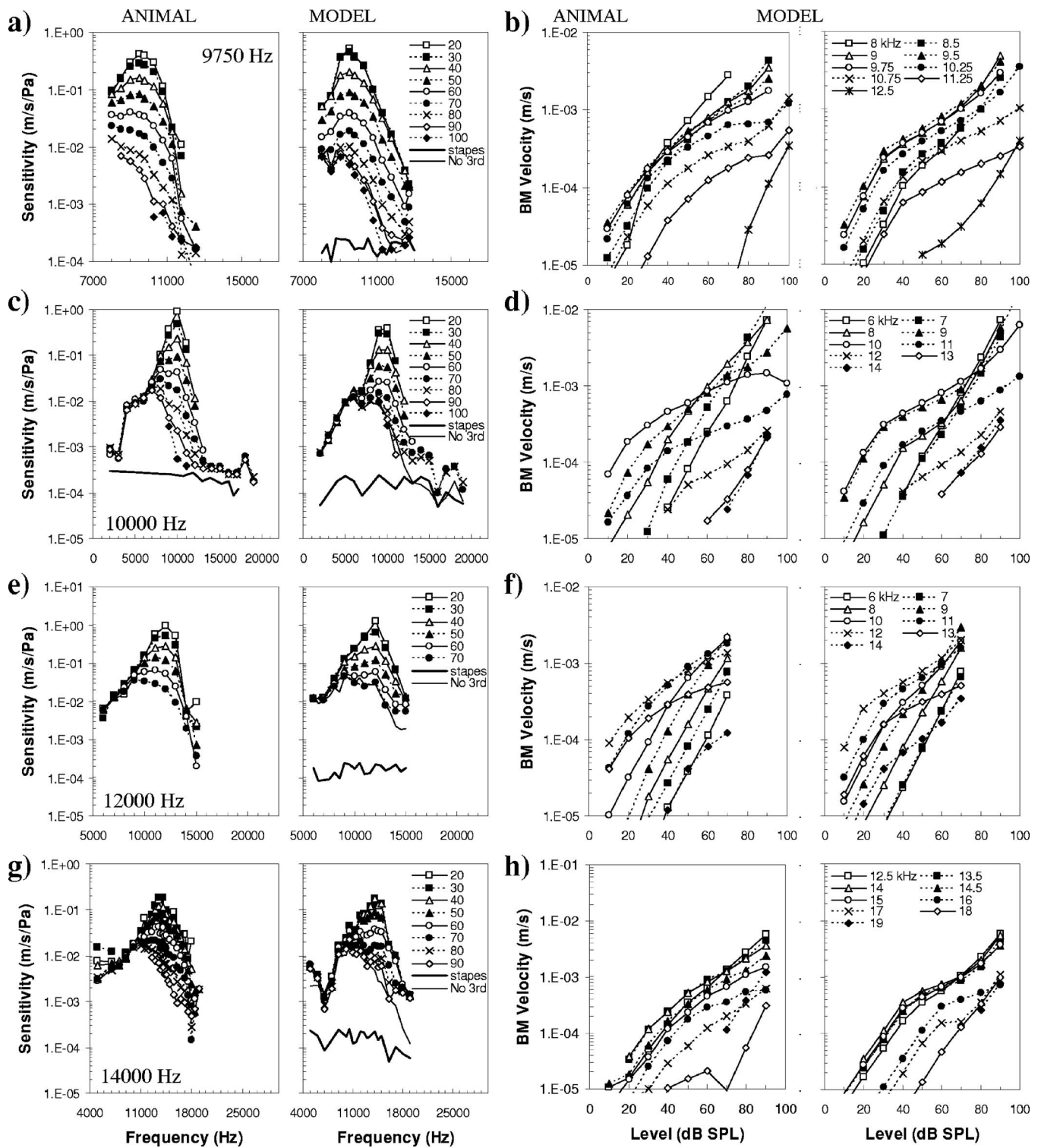
FIG. 4. As for Fig. 3 but for cochlear sites with CFs of 9250, 12 000, 14 000 [data from Rhode and Recio (2000), preparations CB24, CB04, and CB21, respectively], and 10 000 Hz [data from Ruggero *et al.* (1997), animal L113].

qualitatively and quantitatively. The frequency response of the ME filter used in the model exhibits peaks that considerably affect the model frequency response.

7. 14-kHz *site*

The most basal site examined with the model corresponds to case CB21 of Rhode and Recio (2000) and its simulations are shown in the right panels of Figs. 4(g) and

4(h). The model shows a good qualitative and reasonably good quantitative match to the animal data. The experimental I/O curves are significantly compressed for stimulus frequencies above CF; below CF, however, they are almost linear. Model responses to 18- and 19-kHz tones are linear. At 18 kHz, the amplitude of the model response is larger than the experimental one due to the contribution from the third path. At 19 kHz the amplitude of the TRNL-filter response
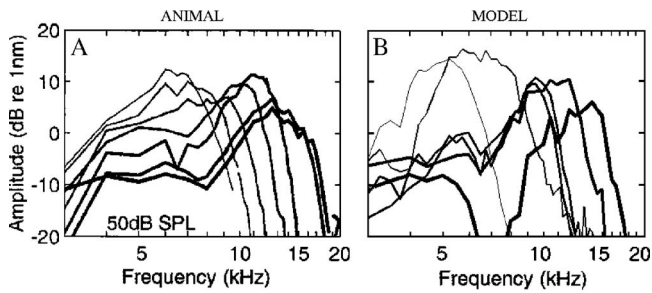
FIG. 5. Comparison of experimental (A) and modeled (B) normalized amplitude responses for a stimulus level of 50 dB SPL. Experimental data are from Rhode and Recio (2000) for CFs from 6 to 14.7 kHz. Model results are for the CFs shown in Table I, except 800 Hz.

matches that of the experimental data. A notch occurs at 7 kHz in the model frequency response [right panel in Fig. 4(g)] that does not occur in the experimental response. This reflects a notch in the ME-filter frequency response and also possibly a phase cancellation between the output signals from the third path and the linear path.

Model amplitude responses normalized to 1 nm for six cochlear sites with CFs between around 5.5 and 12 kHz are plotted in Fig. 5(b) for a sound level of 50 dB SPL. The experimental data [Fig. 5(a)] were for the same sound level and for sites with CFs of 6, 7, 7.9, 10.7, 12.1, 14.3, and 14.7 (Rhode and Recio, 2000). Overall the model frequency responses are comparable to the experimental ones. They show the characteristic broadening at low CFs (notice that the abscissa is on a logarithmic scale). The model slightly underestimates the BW and shows a steeper slope of frequency responses at stimulation frequencies well above CF than the animal data. This is possibly because the number of GT and LP filters in the linear and nonlinear paths were fixed across CFs. Better fits would have been obtained by allowing the number of filters to vary freely across CFs.

B. Phase responses

Phase responses were analyzed at the same time as amplitude responses during parameter optimization. The number of gammatone and low-pass filters and the relation between the center frequencies and the BWs of the gammatone filters in the linear path and the nonlinear path are crucial to the phase response of the model (Lopez-Najera, 2005).

Model and experimental unwrapped phase-frequency responses are compared in Fig. 6. Results are shown only for CFs of 0.8, 7.25, 10, and 12 kHz because those are the only sites for which both phase and amplitude responses have been reported for the same preparation. Figure 6(a) shows animal observations (Rhode and Cooper, 1996, case CH16) and model responses. The model phase advances over one cycle more than in the animal data. This could be corrected by decreasing the number of filters in the linear and nonlinear path, but to reduce the number of degrees of freedom of the system, the number of filters was fixed across CFs. A phase plateau occurs above 1 kHz in the model but at 800 Hz in the data. The model phase is one cycle shorter than in the data at CF=7.25 kHz [animal data in Fig. 6(b) from Rhode and Recio, 2000], and two cycles shorter at CFs
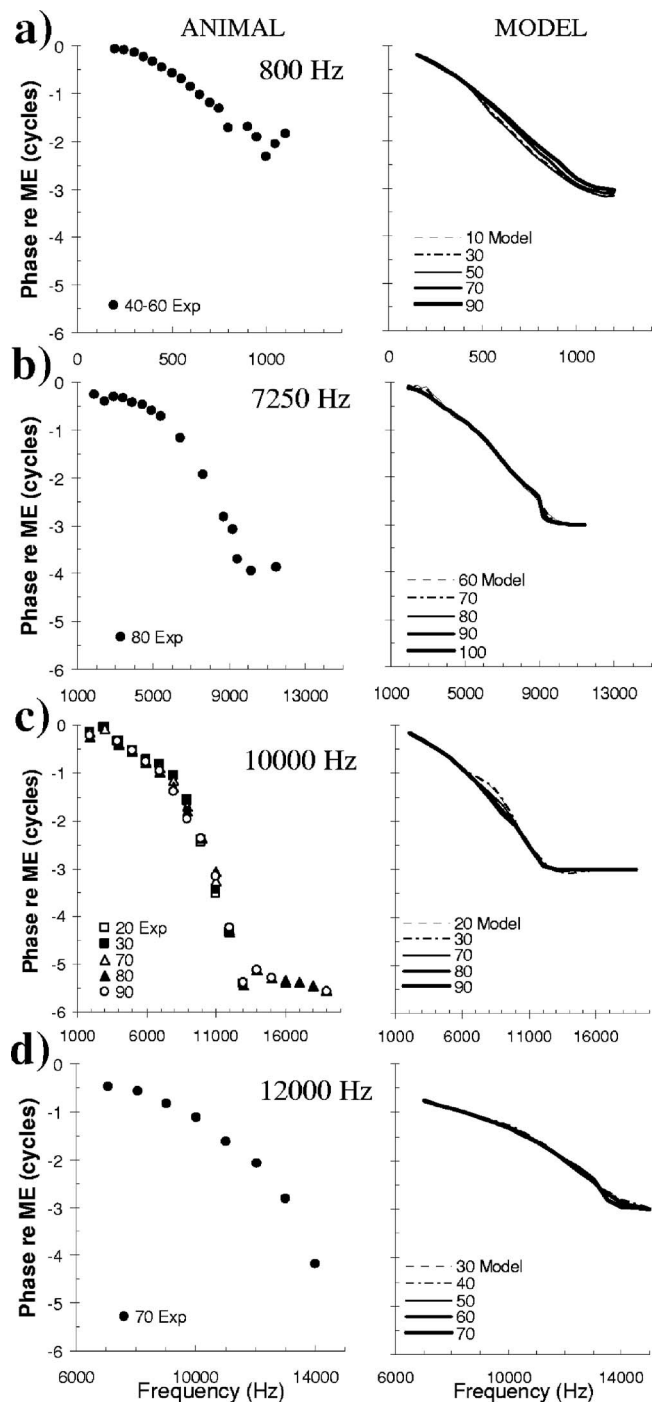
FIG. 6. Experimental and model phase responses relative to the phase of the middle-ear response for CFs of 800 (Rhode and Cooper, 1996), 7250, 12 000 (Rhode and Recio, 2000), and 10 000 Hz (Ruggero et al., 1997). The insets inform of the stimulus level in dB SPL.

of 10 and 12 kHz [animal data in Figs. 6(c) and 6(d) from Ruggero et al. (1997) and Rhode and Recio (2000), respectively].

Figure 7 illustrates the level dependence of model and animal phase responses for CFs of 10 [Fig. 7(a), animal data from Ruggero et al. (1997)] and 7.25 kHz [Fig. 7(c), animal data from Rhode and Recio (2000)]. At 10 kHz, the model simulates qualitative and quantitative aspects of the phase response relative to 80 dB SPL [compare Figs. 7(a) and 7(c)], but only for frequencies below the CF. At CF
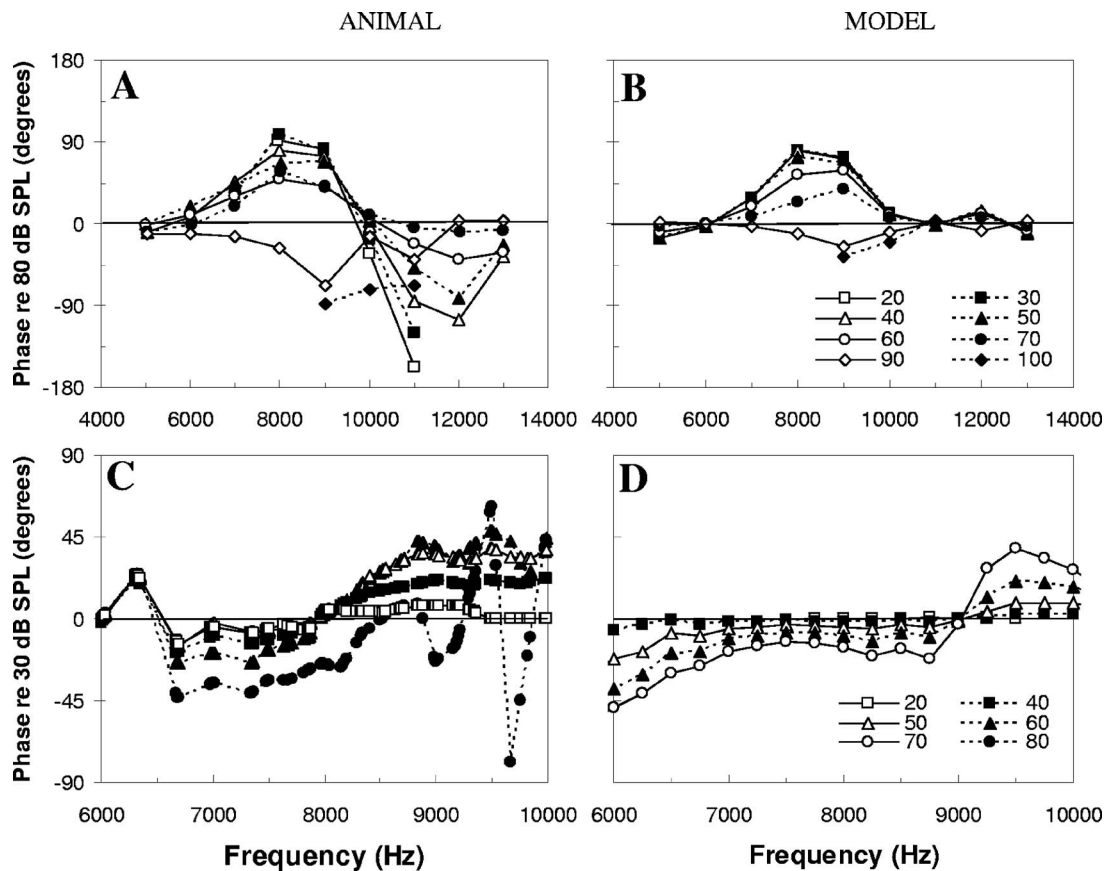
FIG. 7. Experimental and model level-dependent phase responses for cochlear sites with CFs of 10 [(A) and (B)] and 7.25 kHz [(C) and (D)]. The experimental data for 7.25 and 10 kHz are from Rhode and Recio (2000) and Ruggero *et al.*, (1997), respectively. Insets inform of the stimulus level in dB SPL.

=7.25 kHz, the model reproduces qualitatively and (reasonable well) quantitatively the behavior of phase at different level referred to 30 dB SPL. The relation between the center frequency, the BWs, and the number of filters in cascade in the linear and nonlinear paths are critical to reproducing this behavior (Lopez-Najera, 2005).

The longitudinal spatial distribution of BM motion for individual stimulus frequencies has received little attention because the required measurements are difficult. However, measurements exist for the spatial distribution of the excitation produced by a 15-kHz pure tone in the guinea pig cochlea (Russell and Nilsen, 1997) and a complete set of phase responses for several locations along the basal region of the chinchilla cochlea (Rhode and Recio, 2000). Model phase responses are compared with the latter in Fig. 8. Experimental responses are for CFs of 6, 7, 7.9, 10.7, 12.1, and 14.7 kHz, while model responses are for CFs of 5.5, 7.25, 10, 12, and 14 kHz. Model and experimental responses differ
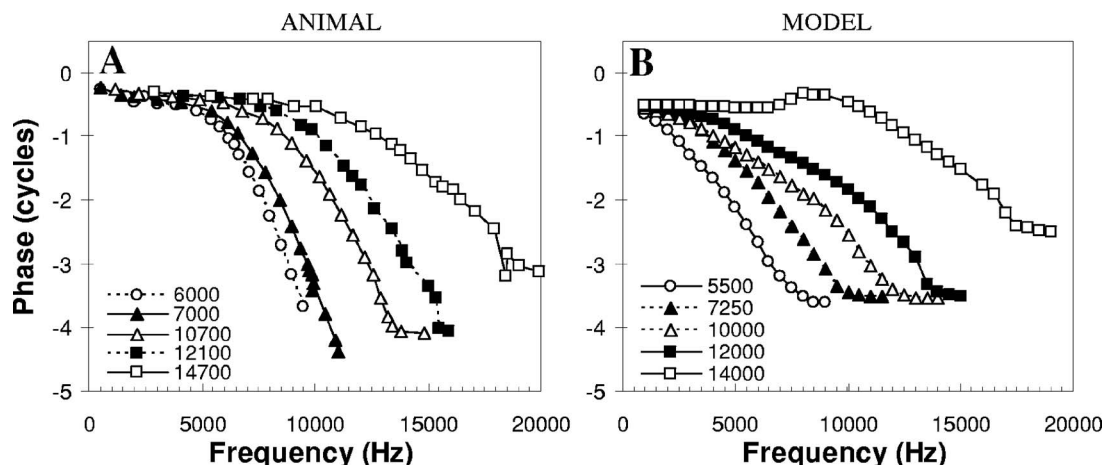
FIG. 8. Experimental and model phase responses at 70 dB SPL for different cochlear sites. Experimental data are from Rhode and Recio (2000). Inset information of the CF in hertz.
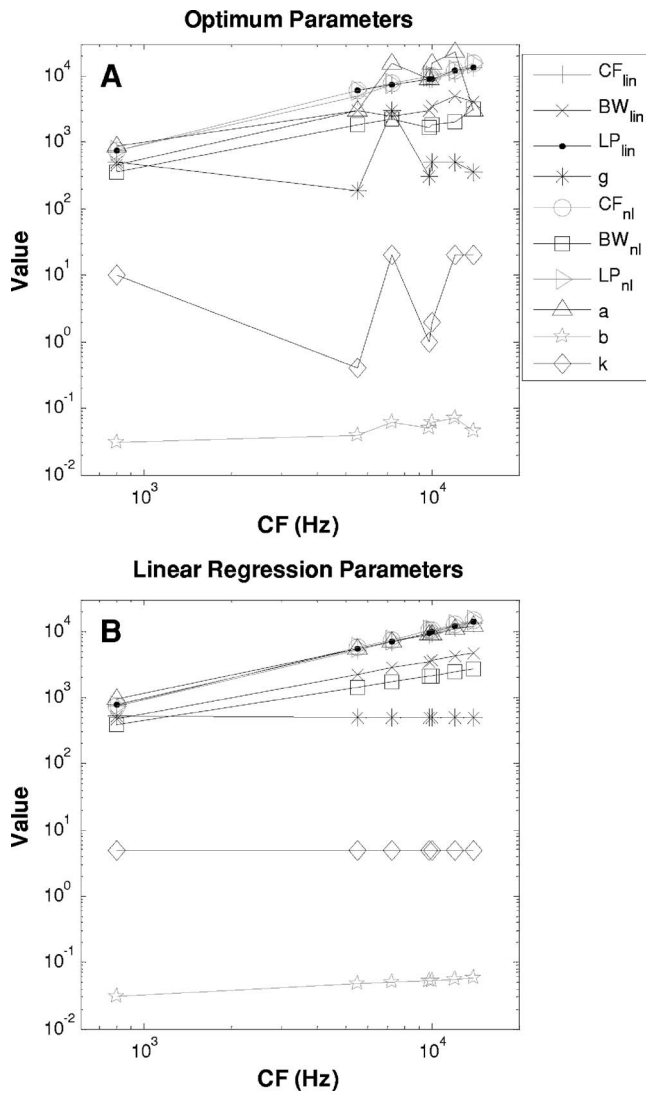
FIG. 9. (A) Values of the optimum model parameters used to model the animal data presented in this report. The actual numerical values are given in Table I. (B) Parameter values as calculated by linear regression of the optimum parameters [Eq. (1)]. Note that some symbols and lines overlap.

slightly, but the main characteristics of the experimental phase responses are simulated reasonably well by the model for CFs from 800 Hz to 14 kHz.

	p_0	m
DRNL filter parameter		
CF_{lin}	−0.01	0.9947
BW_{lin}	0.373	0.7949
LP_{lin}	−0.0187	1.0016
Gain, g	2.7781	−0.0214
CF_{nl}	−0.1545	1.0424
BW_{nl}	0.59	0.6851
LP_{nl}	−0.1545	1.0424
Gain, a	0.3717	0.896
Gain, b	−2.1953	0.2321
k	0.6937	−0.0019

IV. THE FILTERBANK

Having assembled a whole set of parameters for seven different sites along the chinchilla cochlea, the next step was to build a filterbank that allows estimating parameters for intermediate sites. The method used is based on the assumption that the parameters have a linear dependence with CF in a log-log scale [Fig. 9(a)] (cf. Lopez-Poveda and Meddis, 2001). In other words by assuming the following relationship:

$$\log_{10}(\text{parameter}) = p_0 + m \log_{10}(\text{CF}). \qquad (1)$$

The coefficients p_0 and m calculated for the linear regression are shown in Table II. Figure 9(b) shows the new parameters calculated using the filterbank. To be consistent in the analysis, linear regression for all parameters was used, even with k, the gain of the third path, despite the fact that it is less clear that its value depends similarly on CF [Fig. 9(a)].

Figure 10 compares the experimental I/O functions for CF pure tones, for the seven different CFs considered earlier with model simulations using the approximate filterbank parameters. The model captures the fundamental characteristics of the experimental responses. The differences are attributable mainly to the gain values used rather than to the BWs or the filters' center frequencies. It is noteworthy, however, that experimental responses vary more across CF than do model

FIG. 10. Comparison of experimental and model I/O curves for stimuli at CF and for the seven sites considered in this report (see CFs in Table I). Model responses were obtained with the filterbank parameters [see Eq. (1) and Table II] rather than with the optimum parameters.

responses. For instance, the experimental I/O function at 7 kHz is more linear than the I/O curves at any other CFs. These across-CF differences are not captured by the model, but their physiological relevance is uncertain. They might reflect different degrees of cochlear injury across preparations.

V. DISCUSSION

The model presented in this report is able to simulate BM amplitude and phase responses to tones reasonably well. The new TRNL filter improves the match between model and experimental amplitude and phase responses with respect to the DRNL filter of Meddis *et al.* (2001) for frequencies above the CF, in the plateau region. Parameters for seven different sites along the chinchilla cochlea have been reported. A filterbank using seven sets of parameters from 0.8 Hz to 14 kHz has been developed and tested.

An important feature of the present model is its capability of reproducing BM amplitude and phase plateaus that occur for frequencies well above CF. This has been achieved following a suggestion of Robles and Ruggero (2001) that the plateaus reflect more or less directly stapes motion. At one time, consensus was lacking on whether these plateaus were a general feature of BM responses, particularly because they have been hardly reported in corresponding auditory nerve responses. However, plateaus have been widely reported in BM responses of several mammalian species measured using different experimental techniques and in healthy preparations [a comprehensive review of the evidence is provided in pp. 1312–1313 of Robles and Ruggero (2001) and will not be repeated here]. Therefore, they are now accepted as a physiological characteristic of BM responses (Robles and Ruggero, 2001).

The main aim of this study was to test if the DRNL filter was able to simulate BM mechanics from the basal partition of the cochlea to the apical region, and how it was possible, if it was, to improve it. The model responses shown in Figs. 3, 4, and 6 indicate that parameters can be found to reproduce empirical data for different sites over the chinchilla cochlea, even when the number of filters in cascade is previously fixed. The TRNL filter produces gain and phase plateaus for frequencies above BF in each case.

Parameters were optimized by attending not only to I/O functions, but also to amplitude and phase frequency responses, and IRs (not shown). In some instances, the model appears not to capture the experimental phase response in the plateau region. Better fits to the experimental phase responses can be found by changing the number of filters in cascade, but this would introduce another degree of freedom into the system that should be analyzed in further studies (Lopez-Najera, 2005). On the other hand, the experimental phase jumps by more than a cycle between two successive frequencies [e.g., Figs. 6(b)–6(d)] and this casts doubts on the reported experimental phases.

The method for modeling the ME response has improved the model IR (not shown) but it does it at the expense of affecting the frequency response of the model. A simple Butterworth bandpass filter could have been used instead to model the ME transfer characteristics, but it would be unlikely to produce an appropriate phase and IR for the whole system. Some of the discrepancies between the experimental and model results occur because the same ME filter implementation has been used throughout to reproduce BM responses for seven different cochlear sites of seven different animals measured in three different laboratories with different degrees of cochlear injury.

Another weakness of the model [already reported by Meddis *et al.* (2001) and Lopez-Poveda and Meddis (2001)] is that the BF does not change gradually with signal level. Instead, a rather abrupt shift occurs between the BFs of the linear and the nonlinear paths. The main advantage of this model is its simplicity and adding more resonances could improve its responses but it would complicate its simple structure. It is important to remember that the model is a phenomenological model and does not represent cochlear mechanics at a physical level but it is useful to preprocess the input to higher-level auditory models. An approximated transfer function was developed (Lopez-Poveda, 2003) that further simplifies the understanding of the DRNL filter.

[1]The experimental IR of the chinchilla stapes consisted of 1024 samples recorded at sampling rate of 250 kHz (hence the total duration was 0.4096 ms). Since the model was evaluated at a sampling rate of 62.5 kHz, the ME FIR filter coefficients were obtained by downsampling the experimental IR keeping every fourth sample starting with the first one. Hence, the final FIR filter had 256 coefficients.

Cheatham, M. A. and Dallos, P. (**2001**). "Inner hair cell response patterns: Implications for low-frequency hearing," J. Acoust. Soc. Am. **110**, 2034–2044.

Giguere, C., and Woodland, P. C. (**1994**). "A computational model of the auditory periphery for speech and hearing research. I. Ascending path," J. Acoust. Soc. Am. **95**, 331–342.

Goldstein, J. L. (**1990**). "Modeling rapid waveform compression on the basilar membrane as multiple-bandpass-nonlinearity filtering," Hear. Res. **49**, 39–60.

Goldstein, J. L. (**1995**). "Relations among compression, suppression, and combination tones in mechanical responses of the basilar membrane: Data and MBPNL model," Hear. Res. **89**, 52–68.

Händel, P. (**2000**). "Properties of the IEEE-STD-1057 four parameter sine wave fit algorithm," IEEE Trans. Microwave Theory Tech. **49**, 1189–1193.

Irino, T., and Patterson, R. D. (**2001**). "A compressive gammachirp auditory filter for both physiological and psychophysical data," J. Acoust. Soc. Am. **109**, 2008–2022.

Lopez-Najera, A. (**2005**). "Simulación computacional de la membrana basilar" (Computational simulation of the basilar membrane), Ph.D. thesis. Departamento de Biología Celular y Patología. Universidad de Salamanca, Salamanca, Spain.

Lopez-Najera, A., Meddis, R., and Lopez-Poveda, E. A. (**2005**). "A computational algorithm for computing cochlear frequency selectivity: Further studies," in *Auditory signal processing: Physiology, Psychoacoustics and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York), pp. 14–20.

Lopez-Poveda, E. A. (**2003**). "An approximate transfer function for the dual-resonance nonlinear algorithm of auditory frequency selectivity," J.

Acoust. Soc. Am. **114**, 2112–2117.

Lopez-Poveda, E. A. (**2005**). "Spectral processing by the peripheral auditory system: Facts and models," Int. Rev. Neurobiol. **70**, 7–48.

Lopez-Poveda, E. A., and Meddis, R. (**2001**). "A human nonlinear cochlear filterbank," J. Acoust. Soc. Am. **110**, 3107–3118.

Meddis, R., O'Mard, L. P., and Lopez-Poveda, E. A. (**2001**). "A computational algorithm for computing nonlinear auditory frequency selectivity," J. Acoust. Soc. Am. **109**, 2852–2861.

Olson, E. S. (**1998**). "Observing middle and inner ear mechanics with novel intracochlear pressure sensors," J. Acoust. Soc. Am. **103**, 3445–3463.

Oppenheim, A. V., and Schafer, R. W. (**1999**). *Discrete-Time Signal Processing* (Prentice-Hall International, Englewood Cliffs, NJ).

Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). "Basilar-membrane responses to clicks at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **103**, 1972–1989.

Rhode, W. S., and Cooper, N. P. (**1996**). "Nonlinear mechanics in the apical turn of the chinchilla cochlea in vivo," Aud. Neurosci. **3**, 101–121.

Rhode, W. S., and Recio, A. (**2000**). "Study of mechanical motions in the basal region of the chinchilla cochlea," J. Acoust. Soc. Am. **107**, 3317–3332.

Robert, A., and Eriksson, J. L. (**1999**). "A composite model of the auditory periphery for simulating responses to complex sounds," J. Acoust. Soc. Am. **106**, 1852–1864.

Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.

Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **101**, 2151–2163.

Russell, I. J., and Nilsen, K. E. (**1997**). "The location of the cochlear amplifier: Spatial representation of a single tone on the guinea pig basilar membrane," Proc. Natl. Acad. Sci. U.S.A. **94**, 2660–2664.

Sumner, C. J., O'Mard, L. P., Lopez-Poveda, E. A., and Meddis, R. (**2003**). "A nonlinear filter-bank model of the guinea-pig cochlear nerve: Rate responses," J. Acoust. Soc. Am. **113**, 3264–3274.

Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (**2001**). "A phenomenological model for the responses of auditory-nerve fibers. I. Nonlinear tuning with compression and suppression," J. Acoust. Soc. Am. **109**, 648–670.

Zilany, M. S., and Bruce, I. C. (**2006**). "Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery," J. Acoust. Soc. Am. **120**, 1446–1466.

2134   J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Lopez-Najera *et al.*: Modeling basilar membrane frequency

Non-ossicular signal transmission in human middle ears: Experimental assessment of the "acoustic route" with perforated tympanic membranes

Susan E. Voss[a)]
Picker Engineering Program, Smith College, 51 College Lane, Northampton, Massachusetts 01063

John J. Rosowski and Saumil N. Merchant
Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114, Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts 02115 and Speech and Hearing Bioscience and Technology Program, Harvard-M.I.T. Division of Health Sciences and Technology, Cambridge, Massachusetts 02139

William T. Peake
Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, and Speech and Hearing Bioscience and Technology Program, Harvard-M.I.T. Division of Health Sciences and Technology, Cambridge, Massachusetts 02139

Direct acoustic stimulation of the cochlea by the sound-pressure difference between the oval and round windows (called the "acoustic route") has been thought to contribute to hearing in some pathological conditions, along with the normally dominant "ossicular route." To determine the efficacy of this acoustic route and its constituent mechanisms in human ears, sound pressures were measured at three locations in cadaveric temporal bones [with intact and perforated tympanic membranes (TMs)]: (1) in the external ear canal lateral to the TM, P_{TM}; (2) in the tympanic cavity lateral to the oval window, P_{OW}; and (3) near the round window, P_{RW}. Sound transmission via the acoustic route is described by two concatenated processes: (1) coupling of sound pressure from ear canal to middle-ear cavity, $H_{P_{CAV}} \equiv P_{CAV}/P_{TM}$, where P_{CAV} represents the middle-ear cavity pressure, and (2) sound-pressure difference between the windows, $H_{WPD} \equiv (P_{OW} - P_{RW})/P_{CAV}$. Results show that: $H_{P_{CAV}}$ depends on perforation size but not perforation location; H_{WPD} depends on neither perforation size nor location. The results (1) provide a description of the window pressures based on measurements, (2) refute the common otological view that TM perforation location affects the "relative phase of the pressures at the oval and round windows," and (3) show with an intact ossicular chain that acoustic-route transmission is substantially below ossicular-route transmission except for low frequencies with large perforations. Thus, hearing loss from TM perforations results primarily from reduction in sound coupling via the ossicular route. Some features of the frequency dependence of $H_{P_{CAV}}$ and H_{WPD} can be interpreted in terms of a structure-based lumped-element acoustic model of the perforation and middle-ear cavities.
© *2007 Acoustical Society of America.* [DOI: 10.1121/1.2769617]

I. INTRODUCTION

Because the cochlea responds to the pressure difference between the oval and round windows (Voss *et al.*, 1996), it is useful to understand how these "window pressures" are generated in both normal and pathological ears. In normal ears, the net window pressure difference is dominated by the sound pressure delivered to the oval window via the ossicular system, a mechanism referred to as the "ossicular route" by Peake *et al.* (1992). Via the ossicular route, a cochlear window-pressure difference is generated by motion of the tympanic membrane (TM), malleus, incus, and stapes driven by the pressure difference across the TM. In addition to driving the ossicular system, TM motion generates a sound-pressure field in the middle-ear cavity, which for most of the audio frequency range is approximately uniform throughout the cavity (e.g., Lynch, 1981; Voss, 1998). Thus, the sound pressures at the adjacent oval and round windows, P_{OW} and P_{RW}, are nearly equal; the relatively small pressure difference $\Delta P_{win} = P_{OW} - P_{RW}$ adds to the ossicular-route stimulus to the cochlea. Consistent with Peake *et al.* (1992), we refer to the mechanism that drives the cochlea through ΔP_{win} as the "acoustic route" (Fig. 1).

In a normal ear the contribution of the acoustic route is insignificant compared to the ossicular route, but the acoustic route can become important when ossicular transmission is decreased (Békésy, 1947; Schmitt, 1958; Peake *et al.*, 1992; Merchant *et al.*, 1997a). Perforation of the TM decreases

---

[a)]Author to whom correspondence should be addressed. Electronic mail: svoss@email.smith.edu
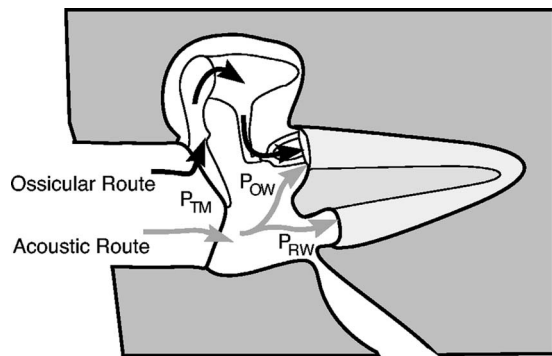
FIG. 1. Schematic representation of two routes for sound transmission to the cochlea. Sound pressure at the TM, P_{TM}, is transmitted to the cochlea through both "ossicular" and "acoustic" routes. In the ossicular route, P_{TM} acts through the coupled motion of the TM, malleus, incus and stapes and produces a pressure $P_{ossicular}$ at the stapes footplate. In the acoustic route, P_{TM} produces a sound-pressure difference $P_{OW} - P_{RW} = \Delta P_{win}$ between the oval and round windows. The net stimulus to the cochlea is the sum of pressures transmitted via the two mechanisms: $P_{ossicular} + \Delta P_{win}$. For low frequencies the acoustic-route stimulus ΔP_{win} is expected to be small relative to P_{OW}, as the oval and round windows are separated by only about 4 mm while the wavelength of sound at 100 Hz is about 3.5 m. In fact, the acoustically coupled stimulus ΔP_{win} is generally only 0.01–0.001 of the ossicularly coupled stimulus. Thus, pathologies that interrupt the ossicular route cause large (40–60 dB) hearing losses. Figure modified from Merchant et al. (1997b).

ossicular-route transmission and purportedly increases acoustic-route transmission. The primary goals of this work are: (1) To describe the acoustic-route transmission systematically and the effect of TM perforations on it; (2) To assess the characteristics of the mechanisms involved in the acoustic route; and (3) To identify conditions in which the acoustic route plays a significant role in hearing.

The effect of TM perforations on the acoustic route has been discussed in the literature, but no systematic measurements have been reported. Békésy (1947) reported measurements of the pressure difference between the oval and round windows at six frequencies in one temporal-bone preparation without a TM. In Peake et al. (1992) these measurements are used with a middle-ear model to predict hearing loss with a total perforation (i.e., TM removed). The model predicts that with the elimination of ossicular coupling, the acoustic route provides hearing levels of 40–60 dB, from 250 to 3000 Hz, consistent with clinically observed hearing loss.

With TM perforations, a popular idea is that a perforation's location in the TM affects the sound pressures at the oval and round windows. Schmitt (1958) introduced the idea that differences in both the magnitude and phase angles of the sound pressures at the two windows are important. Subsequently, a view developed that a perforation directly lateral to the round window affects the pressures acting on the oval and round windows differently from a perforation at other (e.g., more anterior) locations (e.g., Pickles, 1987; pp. 60, 61; Glasscock and Shambaugh, 1990, p. 314; Schuknecht, 1993, p. 196). Specifically, this view assumes that the sound pressure at the round window "cancels" the pressure at the oval window more with posterior-inferior perforations than with perforations at other locations.

Recent work that describes total middle-ear sound transmission in the presence of TM perforations includes: (1) In-

put impedance and ossicular-motion measurements from temporal-bone preparations with controlled, experimentally produced TM perforations (Voss et al., 2001b); (2) An analog-circuit model that describes mathematically the effects of TM perforations on total middle-ear sound transmission (Voss et al., 2001c); (3) Clinical measurements that are consistent with the temporal-bone experimental and theoretical work (Mehta et al., 2006). Collectively, this work describes and explains hearing loss resulting from a TM perforation as:

1. Frequency dependent with the largest loss occurring at the lowest frequencies;
2. Increasing with perforation size;
3. Depending on the volume of the middle-ear and mastoid air space with larger losses in ears with smaller volumes;
4. Not varying significantly with location of the perforation in the TM.

Here we present measurements [from the same preparations reported in Voss et al. (2001b, c)] of sound pressures outside the cochlear windows and describe how these pressures influence sound transmission in the presence of TM perforations. These measurements and analyses place constraints on sound transmission with perforations. Specifically, the importance of the acoustic route relative to the ossicular route is quantified, and the acoustic route's dependences on perforation location, size, and sound frequency are determined.

II. ANALYTICAL APPROACH

To describe sound transmission via the acoustic route, we follow the approach of Békésy (1947) and define the effective acoustic-route (AR) transfer function H_{AR} as the ratio of the pressure difference between the oval and round windows to the pressure in the external ear canal at the TM

$$H_{AR} \equiv \frac{P_{OW} - P_{RW}}{P_{TM}}, \tag{1}$$

where P_{OW} is the pressure within the middle-ear cavity adjacent to the oval window, P_{RW} is the pressure within the middle ear adjacent to the round window, and P_{TM} is the pressure within the ear canal adjacent to the TM. In order to represent the acoustic mechanisms involved (with TM perforations), we express H_{AR} in terms of two factors: the middle-ear-cavity-pressure transfer function $H_{P_{CAV}}$ and the window-pressure-difference transfer function H_{WPD},

$$H_{AR} \equiv \frac{P_{OW} - P_{RW}}{P_{TM}} = \underbrace{\frac{P_{CAV}}{P_{TM}}}_{H_{P_{CAV}}} \underbrace{\frac{P_{OW} - P_{RW}}{P_{CAV}}}_{H_{WPD}} = H_{P_{CAV}} H_{WPD}, \tag{2}$$

where P_{CAV} is the sound pressure within the middle-ear air space. Note that in general P_{CAV} depends on location in the middle-ear cavity; we address this dependence below. The first factor, $H_{P_{CAV}}$, is a measure of the acoustic processes that couple sound pressure in the ear canal to the middle-ear cavity; the second factor, H_{WPD}, is determined by the acoustic processes that produce spatial variations within the cavity. In the absence of a detailed model for the acoustic processes, it

is conceivable that $H_{P_{CAV}}$ and H_{WPD} depend on both location and size of perforations and on frequency.

III. METHODS

A. Transfer function calculations

We use measurements of the oval- and round-window pressures, P_{OW} and P_{RW}, and the ear-canal pressure P_{EC} transformed to pressure at the TM P_{TM} (Voss *et al.*, 2001b), to calculate the factors of the acoustic route. The middle-ear-cavity-pressure transfer function, $H_{P_{CAV}}$, is calculated from our pressure measurements as

$$H_{P_{CAV}} \equiv \frac{P_{CAV}}{P_{TM}} \approx \frac{P_{OW}}{P_{TM}}, \tag{3}$$

where we assume that the cavity pressure P_{CAV} is nearly uniform throughout the cavity and therefore can be approximated in this ratio by the pressure measured near the oval window, P_{OW} (or equally well by P_{RW}). The approximation $P_{CAV} \approx P_{OW}$ is reasonable for frequencies for which the dimensions of the middle-ear cavity are much smaller than the wavelength of sound; i.e., up to about 4000 Hz, where the wavelength (≈ 9 cm) is about ten times the largest cavity dimension. An experimental test of this simplification showed that even near a TM perforation the pressure within the middle-ear cavity is nearly constant in space (Voss, 1998, Figs. 2–7, p. 69).

The window-pressure-difference transfer function H_{WPD} is computed from measurements as

$$H_{WPD} \approx \frac{P_{OW} - P_{RW}}{P_{OW}} = 1 - \frac{P_{RW}}{P_{OW}} \tag{4}$$

The acoustic-route transfer function H_{AR} (Eq. (2)) is the product of the factors $H_{P_{CAV}}$ and H_{WPD}

$$H_{AR} = \underbrace{\frac{P_{OW}}{P_{TM}}}_{H_{P_{CAV}}} \underbrace{\left(1 - \frac{P_{RW}}{P_{OW}}\right)}_{H_{WPD}} \tag{5}$$

B. Experimental overview

Acoustic measurements of middle-ear cavity pressures at the oval and round windows and lateral to the tympanic membrane (TM) were made in cadaveric temporal bones with both intact and perforated TMs. The prepared bones were 11 of the ears for which intact TM results are presented in Voss *et al.* (2000) and for which stapes velocity and impedance at the TM were measured with perforations (Voss *et al.*, 2001a, b); most of the measurement techniques are described in detail in Voss *et al.* (2000, 2001a, b). Additional methods are described here.

Measurements were made on fresh temporal bones for which no evidence of otologic disease was found either in medical records or in oto-microscopic examination. In each case, the bony ear canal was drilled away to expose the TM, and a brass ring was cemented to the bony rim around the TM to allow repeatable coupling and uncoupling of the sound source and calibrated microphone to the ear. Two cali-

brated probe microphones were introduced into the middle-ear cavity with one probe-tube tip adjacent to the oval window and the other to the round window. In each ear, perforations of different sizes and locations were made with an otosurgical Argon laser, as described by Voss *et al.* (2001a, b); generally, a small perforation was made and then perforations of increasing size were centered at the same location [see Fig. 1 of Voss *et al.* (2001b)]. The ear-canal pressure P_{EC} was measured about 3 mm from the TM and these measurements were transformed to approximate the pressure at the TM P_{TM} by modeling the remaining bony ear canal as a uniform cylindrical tube (Voss *et al.*, 2001b, p. 1434).

In these preparations, measurements were also made of stapes velocity and acoustic impedance (Voss *et al.* 2000, 2001a, b, c); some of these results will be useful here in interpretation of the acoustic mechanisms.

C. Measurement of the oval- and round-window pressures

1. Placement of the oval- and round-window probe-tube microphones in the temporal-bone preparation

Probe-tube microphones sensed the sound pressure outside each cochlear window. In each case, the probe-tube microphone consisted of a Knowles hearing aid microphone coupled to a 33-mm-long steel tube with an inner diameter of 0.8 mm and an outer diameter of 1.1 mm; the distal open end of each probe tube was positioned near the appropriate cochlear window. Each probe microphone was attached to the temporal bone as follows. First, a guide tube (inner diameter 1.2 mm), through which the probe tube could slide, was carefully positioned and glued to the temporal bone. The guide tube provided a straight, repeatable path for the probe tube to reach its window. Next, under the view of a dissecting microscope, the open end of the probe tube was inserted through the guide tube until it was as close as possible to its window (0.5–2 mm). Finally, with the position of the probe tube set within the guide tube, a rubber "o-ring" around the probe tube was positioned to "stop" the probe at the proximal end of the guide tube so that the probe tube was held at this location. This arrangement made it convenient to remove the probe-tube microphone assemblies for repeated calibrations throughout an experiment. The distal openings of the probe tubes were at the posterior-most edge of the footplate (oval window) and lateral to the round window.

2. Calibration of the probe-tube microphones

Each probe-tube microphone (one for measurements at the round window and one for measurements at the oval window) was calibrated by coupling it to the sound source and calibrated microphone assembly that was used to generate and measure sound stimuli in the ear canal (Voss *et al.*, 2000); we refer to this calibrated microphone as the "ear-canal microphone" and the calibration procedure to determine the microphone sensitivity in terms of microphone voltage per sound pressure sensed is described in detail by Voss *et al.* (2000). Each of the two window probe-tube mi-

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Voss *et al.*: Non-ossicular signal transmission in human ears   2137

crophones was calibrated relative to the calibrated ear-canal microphone. Together, the tip of each probe-tube microphone was placed within 1 mm of the calibrated ear-canal microphone within a short cylindrical cavity (Voss *et al.*, 2000). Sound stimuli were generated within the cavity and the sound pressure was assumed equal at the ear-canal microphone and the orifice of the probe-tube microphone; the simultaneous responses of each probe-tube microphone and the calibrated ear-canal microphone were used to determine the sensitivity of each probe-tube microphone. This procedure was followed to determine the calibrations for both microphones—specifically—the transfer functions $P_{OW}^{mic}/V_{OW}^{mic}$ and $P_{RW}^{mic}/V_{RW}^{mic}$, which represent the ratios of sound pressure at the oval and round windows to the oval-window and round-window probe-microphone output voltages, respectively.

3. Do the probe-tube microphones measure the pressures at the probe tips?

As the microphones were coupled to the windows via 33-mm-long probe tubes, it is possible that they could sense sound from locations other than the probe tip. To test this possibility, with the microphone coupled to the temporal bone, we measured the microphone output with the probe tube open and also with the window end of the probe tube plugged with cotton. For most frequencies, the output with the probe tube open was at least 40 dB greater (and always at least 20 dB greater) than with the probe tube plugged with cotton (Voss, 1998, Fig. 3-2, p. 112). We concluded that the microphones' outputs are determined by the sound pressure at the probe-tube tip.

4. Errors in relative calibration between the oval- and round-window microphones

The absolute calibration of the oval- and round-window microphones is not critical for our interest in the window-pressure differences, but consistency in the relative calibrations between the oval- and round-window microphones is crucial in determining precisely the pressure difference between the windows

$$P_{OW} - P_{RW} = P_{OW}\left(1 - \frac{P_{RW}}{P_{OW}}\right). \qquad (6)$$

As $|P_{RW}/P_{OW}| \approx 1$, changes in the relative sensitivity of the microphone pair $(P_{RW}^{mic}/V_{RW}^{mic}/P_{OW}^{mic}/V_{OW}^{mic})$, which produce small changes in the inferred P_{OW} and P_{RW}, may have big effects on the inferred pressure difference. However, small changes that are common to the calibration ratios for both microphones will have equally small effects on the pressure difference; these changes are "divided out" in the ratio P_{RW}/P_{OW} and contribute only a small percentage to P_{OW} in Eq. (6) (Voss, 1995).

To detect small changes in the relative calibration (e.g., from moisture in a probe tube), the calibration procedure (Sec. III C 2) was repeated several times during each experiment. We use variation in the relative calibration $P_{OW}^{mic}/V_{OW}^{mic}/P_{RW}^{mic}/V_{RW}^{mic}$, over repeated measurements, to estimate the range of error in the measurement of P_{RW}/P_{OW}.

The process of coupling the probe microphones to the sound-source assembly used for calibrations seemed to be the major source of variability in the calibration. When a probe-tube microphone was attached to the sound-source assembly and left undisturbed, repeated measurements made over several hours showed changes that were less than 0.05 dB in magnitude and 0.001 cycles in angle. In contrast, when the sound-source assembly was moved, and the coupler that attached it to a probe tube was removed and then replaced, the changes in the measured absolute calibrations increased to about 1 dB in magnitude. However, the relative calibration of the two microphones was more stable as it was possible to calibrate both probe-tube microphones, one at a time, without moving the sound source or coupler. Thus, while the absolute calibration of the two microphones changed by 1 dB from calibration to calibration, the relative calibration between the two microphones varied less.

Repeated calibrations among the 11 experiments had different levels of variability. In three of the 11 experiments (bones 23, 22 Right, and 8), the relative calibrations of the oval- and round-window microphones varied by more than 1 dB in magnitude and 0.01 cycles in angle. We considered this level of uncertainty in the calibrations unacceptable for computation of the window-pressure difference (Sec. III C 5). However, in the remaining eight experiments, the relative calibrations of the oval- and round-window microphones varied up to only 0.25 dB[1] in magnitude and 0.005 cycles in angle. Thus, we focus our attention on the results from the subset of eight bones with smaller variability in the ratio between the oval- and round-window microphone calibrations. In one of the eight bones (bone 18), there was a dramatic change (2–3 dB) in the round-window microphone calibration midway through the experiment. It is not clear what happened to the microphone; measurements made after the change are not included here.

5. Inaccuracies in the calculations of P_{RW}/P_{OW}, H_{WPD}, and H_{AR}

Changes in the relative calibration between the microphones limit the accuracy of calculations of H_{AR} and H_{WPD}. In this section, we estimate limits for accurate calculation of these quantities, based on our estimate of ranges of ±0.25 dB in magnitude and ±0.005 cycles in angle.

First consider how a variation of ±0.25 dB in magnitude and ±0.005 cycles in angle affects the ratio P_{RW}/P_{OW} when $P_{RW} \approx P_{OW}$. A 0.25 dB change in magnitude in either P_{OW} or P_{RW} corresponds to a factor of 1.03. Thus, if the pressures are equal (i.e., $P_{RW}=P_{OW}$), the measurement of the magnitude $|P_{RW}/P_{OW}|$ could range from 0.97 to 1.03, and the angle measurement could range from −0.005 to 0.005 cycles when the actual angle $\angle(P_{RW}/P_{OW})=0$. Next, consider how a relative variation of ±0.25 dB in magnitude and ±0.005 cycles in angle affects the measured $H_{WPD}=(1-P_{RW}/P_{OW})$ (Eq. (4)), when $P_{RW}=P_{OW}$. With $P_{RW}/P_{OW}=1.03\angle0.005$, $|H_{WPD}|$ is 0.04 instead of zero. Small errors in magnitude can introduce variation in the angle $\angle(H_{WPD})$ of ±0.5 cycles because the calculation of H_{WPD} depends on a difference that can have a

2138   J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Voss *et al.*: Non-ossicular signal transmission in human ears

TABLE I. Measurement uncertainties or edge of inaccuracy for calculation of P_{RW}/P_{OW}, H_{WPD} and H_{AR}, based on the assumption that relative errors in the measurements of the oval- and round-window sound pressures are less than ±0.25 dB in magnitude and ±0.005 cycles in angle.

Calculated quantity	Measurement uncertainty or edge of inaccuracy
$\|P_{RW}/P_{OW}\|$	Uncertainty ±0.03
$\angle(P_{RW}/P_{OW})$	Uncertainty ±0.005 cycles
$\|H_{WPD}\|$	Edge of inaccuracy: 0.04
$\angle(H_{WPD})$	Uncertainty ±0.50 cycles
$\|H_{AR}\|$	Edge of inaccuracy: $0.04\left\|\dfrac{P_{OW}}{P_{TM}}\right\|$
$\angle H_{AR}$	Uncertainty ±0.50 cycles

negative or positive sign depending on the relative values of $\|P_{RW}\|$ and $\|P_{OW}\|$; thus, we have chosen to ignore $\angle(H_{WPD})$ because of possibly large errors.

Finally, consider how the relative changes of ±0.25 dB and ±0.005 cycles in P_{RW} and P_{OW} affect the acoustic-route transfer function H_{AR}. If we assume that measurement errors in the ratio P_{OW}/P_{TM} (i.e., $H_{P_{CAV}}$) are negligible, then (from Eq. (5)), errors in $\|H_{AR}\|$ are equal to errors in $\|H_{WPD}\|$ multiplied by the factor $\|P_{OW}/P_{TM}\|$. Thus, when the calculated $\|H_{AR}\| < 0.04\|P_{OW}/P_{TM}\|$, H_{AR} may be strongly influenced by relative microphone calibration changes. We define a boundary at $\|H_{AR}\| = 0.04\|P_{OW}/P_{TM}\|$ and refer to this as the "edge of inaccuracy in the measurement of $\|H_{AR}\|$." As discussed above in relation to $\angle(H_{WPD})$, when $\|H_{AR}\| < 0.04\|P_{OW}/P_{TM}\|$ our measurements of angle $\angle H_{AR}$ contain uncertainties of up to ±0.5 cycles.

Table I summarizes the uncertainties and their influence on our ability to measure the quantities P_{RW}/P_{OW}, H_{WPD} and H_{AR}. In the results that follow, we plot all of our magnitude measurements, but we indicate the edge of inaccuracy based on Table I.

D. Statistical tests to compare two groups of data

We use a statistical test to determine the significance of differences between different populations within our data. Specifically, to test the null hypothesis that two means are from the same population, we compute the probability (p) that a given pair of means could differ by as much as the data show, given the null hypothesis. In particular, we test to determine the p values associated with the means for $\|P_{RW}/P_{OW}\|$, $\angle(P_{RW}/P_{OW})$, $\|H_{WPD}\|$, and $\|H_{AR}\|$ when there are (1) different sized perforations and (2) perforations at different TM locations.

Since the sample sizes used to compute the means are small (range $N=2$ to $N=16$), we compute the corresponding p values using a bootstrap permutation test (Efron and Tibshirani, 1993); this method makes no assumption about the probability distribution. The probability that the two means are equal depends on the fraction of the permutations in which the difference between the means of the two permutated test populations are greater than or equal to the difference between the actual means. Further details of our application of this permutation test can be found in Voss (1998).

IV. RESULTS

A. Organization

Each of three sections presents results from our measurements of sound transmission via the acoustic route: (1) the middle-ear-cavity transfer function $H_{P_{CAV}} = P_{OW}/P_{TM}$; (2) the ratio between the oval- and round-window pressures P_{RW}/P_{OW} and the window-pressure-difference transfer function magnitude $\|H_{WPD}\| = \|1 - P_{RW}/P_{OW}\|$; (3) the acoustic-route transfer function $H_{AR} = \|(P_{OW} - P_{RW})/P_{TM}\|$. For each of these the effects of both perforation size and location are examined. In general, we present results from one of our preparations (bone 24L) as an example and mean measurements across all preparations when appropriate. [Results from each preparation are in Voss (1998).]

B. Middle-ear-cavity-pressure transfer function: $H_{P_{CAV}}$

1. $H_{P_{CAV}}$ measurements with an intact tympanic membrane

Measured middle-ear-cavity-pressure transfer functions are presented in Fig. 2 with TM perforation sizes varying from zero (intact) to total (TM removed) for the example bone (24L) and for mean measurements from 11 ears. Both the individual measurements (on bone 24L) and the mean measurements are representative of all measurements of $H_{P_{CAV}}$ on the 11 preparations.

With the TM intact, at low frequencies (i.e., <1000 Hz) $\|H_{P_{CAV}}\|$ is roughly independent of frequency at about -10 ± 1 dB on bone 24L (Fig. 2) and across the 11 preparations ranges from -6.6 to -21.1 dB at 488 Hz (mean -12.7 dB and standard deviation 4.8 dB). In the 1000–3000 Hz range, $\|H_{P_{CAV}}\|$ has a sharp minimum: across the 11 preparations this magnitude minimum ranges from -20.5 to -39.2 dB (mean -30.5 dB and standard deviation 6 dB) and the frequency of the minimum ranges from 928 to 2686 Hz (mean 2151 Hz and standard deviation 531 Hz). In some preparations, at a lower frequency, a peak in magnitude occurs such as that near 800 Hz in Fig. 2 (upper, left); in other preparations there is no peak in this region. In all preparations, there is a sharp maximum in $\|H_{P_{CAV}}\|$ at a frequency above the sharp minimum described above; across the 11 preparations this maximum occurs between 2612 and 4785 Hz (mean 3650 Hz and standard deviation 564 Hz) and has a value that ranges from -10.3 to -0.8 dB (mean -4.5 dB and standard deviation 2.7 dB).
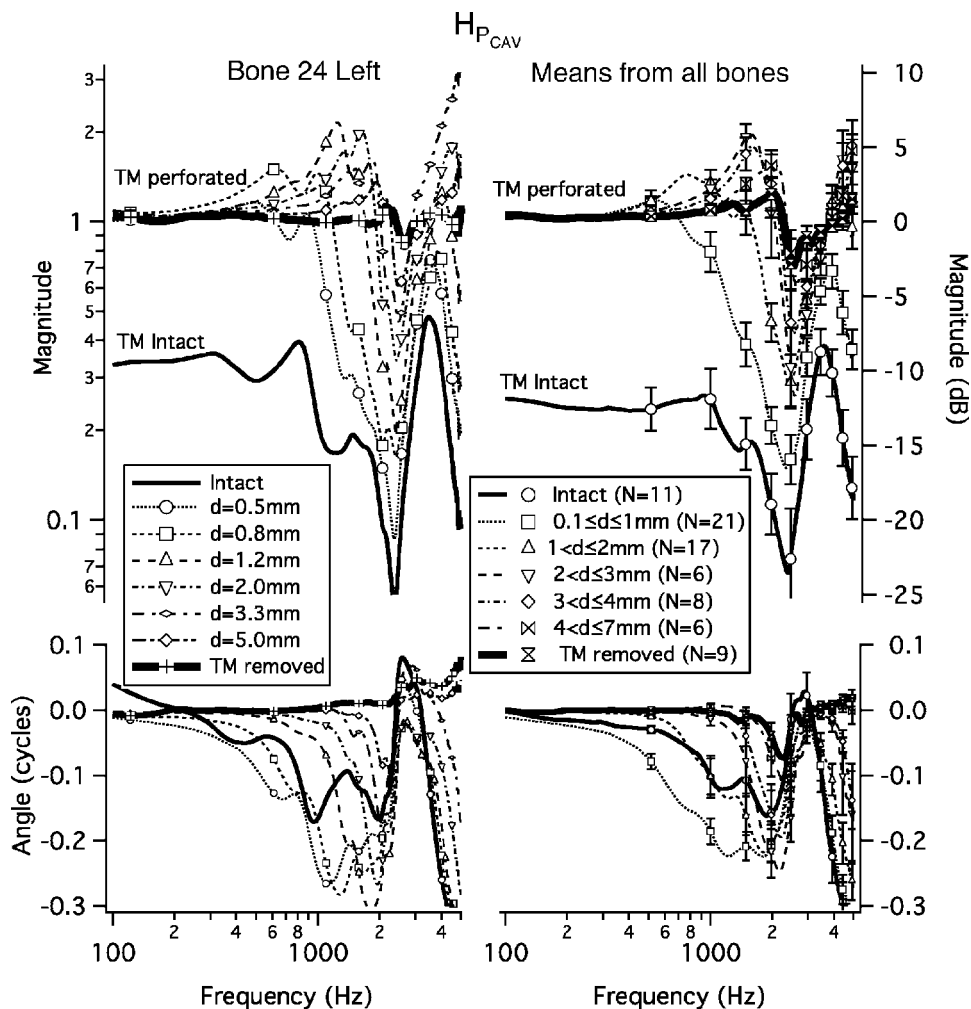
FIG. 2. Measurements of the middle-ear-cavity pressure transfer function $H_{P_{CAV}} = P_{OW}/P_{TM}$ with an intact TM and with perforations of increasing diameter. Symbols are spaced by 20 data points. (There are no symbols for the intact case.) Left panel: Bone 24L. Right panel: Means and standard errors from all measurements. Upper: Magnitude. The two vertical coordinate systems are equivalent; on the left "Magnitude," is the $|H_{P_{CAV}}|$ and on the right "Magnitude(dB)" $= 20\log_{10}$ "Magnitude." Lower: Angle.

The angle $\angle H_{P_{CAV}}$ is about zero at the lowest frequencies (<200 Hz) with the TM intact,[2] indicating that the cavity and ear-canal pressures are in phase. As frequency increases toward the frequency of the minimum in $|H_{P_{CAV}}|$, $\angle H_{P_{CAV}}$ decreases so that the middle-ear cavity pressure lags the ear-canal pressure by up to 0.3 cycles. At the frequency of the sharp minimum in $|H_{P_{CAV}}|$, $\angle H_{P_{CAV}}$ increases rapidly toward a value between 0 and 0.1 cycles. As frequency increases further, $\angle H_{P_{CAV}}$ decreases rapidly in the 3000–4000 Hz range of the local maximum in $|H_{P_{CAV}}|$.

In summary, with the TM intact, for frequencies from 100 to 5000 Hz, $H_{P_{CAV}}$ shows behavior consistent with a system with at least two resonances, represented by a mid-frequency sharp minimum in magnitude and a higher frequency sharp maximum in magnitude.

2. $H_{P_{CAV}}$ with a tympanic-membrane perforation

In additional to the plots in Fig. 2, the features described here of $H_{P_{CAV}}$ with a perforated TM are summarized in Table II for all 11 preparations and all perforation sizes. At the lowest frequencies $H_{P_{CAV}}$ (with a TM perforation) is approximately one (i.e., magnitude=0 dB and angle=0) (Fig. 2);

physically, the perforation acts as an acoustic short circuit between the ear canal and middle-ear cavity, making the pressures equal. As frequency increases above 300 Hz, $|H_{P_{CAV}}|$ peaks above 0 dB (i.e., the middle-ear cavity pressure is greater than the ear-canal pressure); the frequency of this peak increases systematically with perforation diameter. Above the frequency of the peak, $|H_{P_{CAV}}|$ has a sharp minimum (near 2000 Hz); the frequency of this minimum appears to be the same as seen with the TM intact, independent of perforation diameter; $|H_{P_{CAV}}|$ at the minimum increases with perforation size increases. At higher frequencies, the location of the sharp maximum in $|H_{P_{CAV}}|$ that occurs with the TM intact increases somewhat in frequency and magnitude with increasing perforation diameter.

The angle $\angle H_{P_{CAV}}$ (with the TM perforated) is near zero at the lowest frequencies. As frequency increases, $\angle H_{P_{CAV}}$ decreases; the frequency at which the angle decrease begins increases as perforation diameter increases. For both the intact and perforated conditions, $\angle H_{P_{CAV}}$ increases rapidly at the frequency for which $|H_{P_{CAV}}|$ is a minimum, and $\angle H_{P_{CAV}}$ again decreases rapidly near the frequency of the local maximum in $|H_{P_{CAV}}|$.

TABLE II. Features of $H_{P_{CAV}}$ discussed in the text for each preparation. The columns labeled "Bone" and "Perf. Diam" indicate the preparation number and the state of the TM as intact, removed, or by the diameter of the perforation; the column labeled $|H_{P_{CAV}}|$ Max(0.5–2.8 kHz) indicates the magnitude and frequency location of the local maximum that occurs between 0.5 and 2.8 kHz; the column labeled $|H_{P_{CAV}}|$Min (0.9–3.3 kHz) indicates the magnitude and frequency location of the local minimum that occurs between 0.9 and 3.2 kHz; the column labeled $|H_{P_{CAV}}|$ Max($>$2.5 kHz) indicates the magnitude and frequency location of the local maximum at a frequency greater than 2.5 kHz; the column labeled $\angle H_{P_{CAV}}$ (0.2–2.7 kHz) indicates the lowest frequency at which the angle $\angle H_{P_{CAV}}$ decreases from greater than to less than −0.03 cycles; the column labeled $|Z_{TM}|$Min (0.5–3.0 kHz) indicates the frequency of the local minimum that occurs between 0.5 and 3.0 kHz in the magnitude of the impedance measured in the ear canal at the tympanic membrane. Entries for which the feature does not exist in the data are indicated by DNE (does not exist). The entry "noisy" indicates the measurement was too noisy to identify the extremum. The entry "NA" indicates the measurement was not available.

| | | Middle-ear cavity transfer function $H_{P_{CAV}}(f,d)$: Measures of salient features | | | | | | | |
| | Preparation | $|H_{P_{CAV}}|$ Max 0.5–2.8 kHz | | $|H_{P_{CAV}}|$ Min 0.8–3.2 kHz | | $|H_{P_{CAV}}|$ Max $>$2.5 kHz | | $\angle H_{P_{CAV}}$ 0.2–2.7 kHz | $|Z_{TM}|$ Min 0.5–3.0 kHz |
Bone No.	Perf. Diam. (mm)	Mag. (dB)	f_{max} (kHz)	Mag. (dB)	f_{min} (kHz)	Mag. (dB)	f_{max} (kHz)	lowest f for $\angle = -0.03$ (kHz)	f_{min} (kHz)
	Intact	DNE	DNE	−39.2	2.39	−6.0	4.79	DNE	0.98
	0.5	−1.5	1.00	−34.0	2.32	−2.5	5.37	DNE	1.03
	0.7	1.8	1.05	−30.3	2.39	0.2	4.93	0.37	1.07
8	1.1	7.40	1.12	−28.9	2.39	5.3	5.96	0.59	1.17
	1.6	8.5	1.27	−23.2	2.39	9.4	6.15	0.66	1.34
	2.4	9.8	1.39	−21.7	2.39	11.4	6.54	0.76	1.44
	3.2	12.3	1.56	−21.9	2.37	11.7	7.28	0.98	1.61
	Removed	10.3	2.00	−15.5	2.47	11.1	8.33	1.10	1.96
	Intact	DNE	DNE	−26.1	0.93	−0.8	2.61	DNE	0.98
	0.1	DNE	DNE	−23.5	0.85	0.2	2.56	DNE	0.93
9	1.1	6.3	0.54	−17.4	0.81	7.2	2.91	0.39	0.56
	1.9	6.3	0.71	−11.9	0.93	7.1	3.39	0.59	0.73
	3.3	4.4	0.90	−7.7	1.17	4.5	4.00	0.76	0.98
	Removed	DNE	DNE	−0.2	1.00	1.4	4.32	DNE	0.98
	Intact	DNE	DNE	−34.4	1.37	−2.0	3.44	DNE	1.86
	0.5	DNE	DNE	−28.5	1.32	−3.0	3.44	DNE	0.73
13	0.8	6.9	0.61	−24.0	1.37	−1.7	3.56	0.39	0.73
	1.4	7.9	0.76	−23.4	1.25	3.7	3.59	0.56	0.81
	1.8	9.3	0.90	−21.1	1.29	3.7	3.98	0.71	0.98
	Removed	3.2	1.27	−3.2	1.49	DNE	DNE	1.22	1.37
	Intact	DNE	DNE	−35.4	2.44	−10.3	3.22	DNE	NA
	0.4	DNE	DNE	−25.4	2.25	−7.2	3.17	DNE	NA
	0.5	3.6	0.59	−22.1	2.25	−3.6	3.20	0.24	NA
18	1.0	7.8	0.88	−16.6	2.22	0.0	3.37	0.59	NA
	1.7	6.6	1.10	−15.3	2.22	7.5	3.59	0.81	NA
	2.3	5.8	1.20	−9.4	2.29	8.0	3.71	0.93	NA
	3.0	5.1	1.34	−7.0	2.37	10.3	4.00	1.15	NA
	Removed	4.2	2.12	0.0	2.69	16.3	4.98	2.00	NA
	Intact	DNE	DNE	−25.8	2.10	−4.1	3.44	DNE	2.49
	0.6	4.5	0.85	−16.2	2.25	1.0	3.76	0.59	0.88
19	1.3	6.4	1.07	−12.2	1.95	5.5	3.96	0.83	1.10
	1.9	5.6	1.34	−10.2	1.98	7.9	4.15	1.00	1.42
	2.6	6.2	1.44	−8.2	2.03	10.8	4.44	1.12	1.46
	Removed	4.7	1.76	−2.8	2.29	31.2	6.01	1.73	1.98
	Intact	DNE	DNE	−28.2	2.15	−3.7	4.08	DNE	1.90
	0.7	4.6	0.71	−19.6	2.10	2.9	4.27	0.46	0.81
	0.7	4.4	0.66	−19.1	2.10	2.7	4.22	0.42	0.73
	1.0	5.4	0.83	−13.2	2.69	4.4	4.37	0.63	0.85
20	1.7	5.5	1.17	−9.9	2.59	8.2	4.71	0.90	1.59
	3.0	6.3	1.56	−8.1	2.64	12.6	5.42	1.15	1.61
	4.0	5.5	1.64	−6.4	2.61	17.2	5.93	1.29	1.73
	5.0	4.2	1.64	−5.9	2.66	20.0	6.13	1.37	1.81
	Removed	2.6	2.27	−2.2	2.66	29.5	7.47	2.22	2.22

TABLE II. (*Continued.*)

		$|H_{P_{CAV}}|$ Max 0.5–2.8 kHz		$|H_{P_{CAV}}|$ Min 0.8–3.2 kHz		$|H_{P_{CAV}}|$ Max >2.5 kHz		$\angle H_{P_{CAV}}$ 0.2–2.7 kHz	$|Z_{TM}|$ Min 0.5–3.0 kHz
Bone No.	Preparation Perf. Diam. (mm)	Mag. (dB)	f_{max} (kHz)	Mag. (dB)	f_{min} (kHz)	Mag. (dB)	f_{max} (kHz)	lowest f for $\angle=-0.03$ (kHz)	f_{min} (kHz)
22L	Intact	DNE	DNE	−20.5	2.32	−2.8	3.59	DNE	2.29
	0.6	1.3	0.93	−17.5	2.15	1.7	3.74	0.24	0.93
	1.2	5.4	1.05	−15.3	2.20	5.6	4.03	0.61	1.10
	2.0	7.1	1.37	−10.8	2.34	7.5	4.25	1.17	1.59
	3.3	7.8	1.71	−7.9	2.39	8.7	4.61	1.44	1.81
	5.0	7.0	1.86	−5.1	2.39	11.2	5.35	1.73	1.98
	Removed	4.5	2.12	−3.0	2.61	9.9	5.91	2.03	2.20
22R	Intact	DNE	DNE	−29.0	2.69	−5.8	3.66	DNE	1.76
	0.6	0.9	1.27	−25.5	2.69	−1.0	3.76	0.37	1.27
	1.3	7.7	1.46	−22.7	2.69	6.8	4.03	0.76	1.47
	2.3	11.8	1.68	−18.4	2.69	11.4	4.37	1.34	1.76
	3.3	12.1	1.95	−14.4	2.69	15.1	4.96	1.73	2.03
	6.8	8.6	2.20	−9.8	2.81	Noisy	Noisy	2.10	2.37
	Removed	2.4	2.20	−4.7	2.83	Noisy	Noisy	1.15	2.37
23	Intact	DNE	DNE	−29.0	2.69	−5.1	3.59	DNE	1.88
	0.8	DNE	DNE	−25.5	2.69	−0.4	3.71	0.29	1.71
	1.0	2.6	1.51	−22.7	2.69	1.6	3.83	0.56	1.73
	2.0	8.6	1.83	−18.4	2.69	6.1	4.03	0.98	1.86
	3.0	10.3	2.03	−14.4	2.69	7.5	4.27	1.44	2.03
	4.0	10.4	2.20	−9.8	2.81	7.8	4.54	1.73	2.15
	5.0	5.7	2.37	−4.7	2.83	Noisy	Noisy	2.27	2.54
	Removed	DNE	DNE	−2.3	3.15	Noisy	Noisy	DNE	2.81
24L	Intact	DNE	DNE	−25.7	2.37	−6.4	3.49	DNE	1.86
	0.5	0.6	0.88	−21.3	2.37	−2.5	3.52	0.27	0.95
	0.8	3.3	1.00	−15.8	2.39	−2.0	3.83	0.46	1.03
	1.2	6.7	1.25	−14.2	2.42	2.0	4.10	0.78	1.29
	2.0	6.1	1.64	−9.3	2.42	5.4	4.47	1.20	1.59
	3.3	4.1	1.83	−6.7	2.49	10.6	4.98	1.44	1.98
	5.0	2.7	1.96	−4.3	2.47	15.5	6.18	1.89	1.98
	Removed	1.1	2.17	−1.7	2.64	22.3	7.67	DNE	2.10
24R	Intact	DNE	DNE	−32.8	2.54	−2.2	4.22	DNE	2.88
	0.5	2.6	0.73	−30.7	2.61	0.3	4.22	0.37	0.83
	0.6	2.3	0.73	−27.7	2.61	2.0	4.13	0.37	0.78
	1.3	5.5	1.44	−19.5	2.54	7.2	4.56	0.98	1.54
	2.3	8.5	1.81	−14.0	2.61	10.9	5.20	1.71	1.81
	4.0	10.0	2.10	−11.3	2.61	11.0	5.79	1.73	2.10
	5.0	8.7	2.25	−7.1	2.61	15.0	6.37	1.98	2.22
	Removed	4.1	2.79	−1.3	2.49	21.9	8.03	2.71	2.61

In summary, perforations have dramatic effects on the middle-ear-cavity transfer function $H_{P_{CAV}}$ and increase its magnitude relative to the intact case. For the lowest frequencies, all perforation sizes used equalize the ear-canal and middle-ear-cavity pressures so that $H_{P_{CAV}} \approx 1$. Perforations introduce a resonance at which the cavity pressure magnitude exceeds the ear-canal pressure magnitude by up to 10 dB; the resulting local maximum in $|H_{P_{CAV}}|$ occurs between 0.5 and 2.8 kHz, and in each preparation the frequency of the maximum increases as perforation diameter increases. A minimum in $|H_{P_{CAV}}|$ occurs above the frequency of this first peak (in the 1000–3000 Hz range); the frequency of the minimum for each preparation appears to be independent of the TM condition (i.e., intact, various perforation sizes, or TM removed), and the minimum $|H_{P_{CAV}}|$ increases with increasing perforation diameter. A second peak occurs at a frequency above the frequency of this minimum, and the frequency of this peak increases with increasing perforation diameter. These trends occur in each preparation (Table II).

3. Effects of perforation location on $H_{P_{CAV}}$

To test for effects of perforation location on $H_{P_{CAV}}$, we measured $H_{P_{CAV}}$ in two preparations [discussed in Voss *et al.*
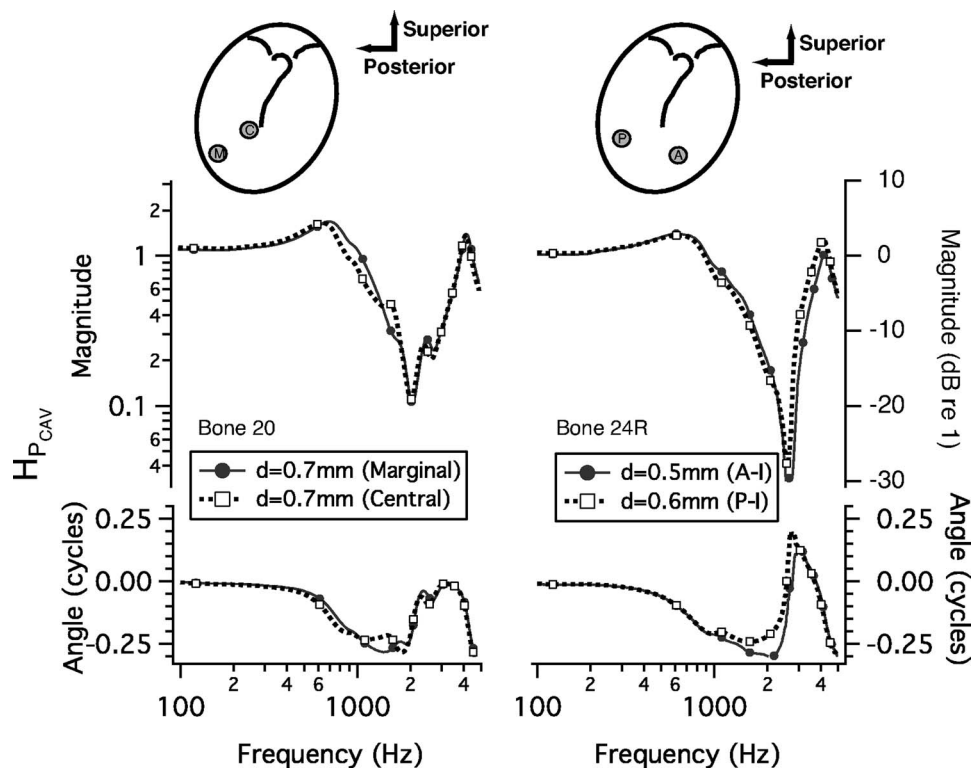
FIG. 3. Comparisons of $H_{P_{CAV}}$ with a perforation at one of two locations. Cartoons at the top portray the TM with manubrium of the malleus as seen from the external ear canal. Patches made of cigarette paper were applied to close a perforation of the TM so that measurements represent the effect of only one perforation. Measurements (not shown here) with a patch in place demonstrate that the response with the patch was similar to the response with the TM intact (Voss *et al.*, 2001a). Left panel: A perforation in either the marginal (M) or central (C) part of the posterior-inferior quadrant. Right panel: A perforation in either the posterior-inferior (P) or anterior-inferior (A) quadrants. Symbols are at every 20th data point. Upper: Magnitude. The vertical coordinate systems are equivalent, with the scale on the left the pressure ratio and the scale on the right the ratio in dB. Lower: Angle.

(2001a) (bones 20, 24 Right)], in each of which measurements were made with a perforation at either of two locations. Voss *et al.* (2001a) demonstrated that patching small perforations (diameter less than 1 mm) with cigarette paper (moistened with saline) returns the middle-ear response (stapes velocity) to that with an intact TM, thus allowing comparison of responses with one perforation size at different locations.

Figure 3 compares the effects of perforations at two locations on $H_{P_{CAV}}$ for these two cases. In each case, the difference in $H_{P_{CAV}}$ (magnitude and angle) between the two perforation locations is small compared to the large variation with frequency; thus, $H_{P_{CAV}}$ appears to be independent of a perforation's location for the comparisons (a) marginal versus central (Fig. 3, left column) and (b) anterior-inferior versus posterior-inferior (Fig. 3, right column). The slight differences in $H_{P_{CAV}}$ in the two locations could result from small differences in the perforation size or changes in $H_{P_{CAV}}$ from intact to patched TM. These results extend the conclusion of Voss *et al.* (2001a) that the effects of a TM perforation on the middle-ear's response to sound do *not* depend on the perforation's location in the TM.

C. Window-pressure-difference transfer function: H_{WPD}

Measurements of the pressure ratio P_{RW}/P_{OW} and the calculated transfer function $H_{WPD}=1-P_{RW}/P_{OW}$ were analyzed for the subset of eight ears in which the probe-tube

microphone calibrations were least variable (Sec. III C 4). Figure 4 plots P_{RW}/P_{OW} and $|H_{WPD}|=|1-P_{RW}/P_{OW}|$ from one preparation (bone 24L); these plots are representative of the results from the other seven bones, as described below. Mean results across ears are not displayed, as inter-ear differences in the frequencies of magnitude maxima and minima result in means that are not representative of the shape of individual measurements; instead, features of these transfer functions in the eight preparations are presented in Table III.

1. Frequency dependence

In all eight preparations, at the lower frequencies (i.e., less than 1000 Hz), the magnitude $|P_{RW}/P_{OW}| \approx 1$ and the angle $\angle(P_{RW}/P_{OW}) \approx 0$, indicating that $P_{RW} \approx P_{OW}$; these features appear for both the intact TM cases and all perforation sizes, and are apparent in the example bone 24L in Fig. 4 (left). Above 1000 Hz, a frequency region exists where $|P_{RW}/P_{OW}|$ exhibits a minimum with a magnitude less than 1 and at a higher frequency a maximum with a magnitude greater than 1; for example, in bone 24L with an intact TM, the magnitude minimum is about 0.7 at 2200 Hz and the magnitude maximum is about 1.2 at 2500 Hz (Fig. 4, left). In the frequency range of these extrema there is a sharp increase in $\angle(P_{RW}/P_{OW})$ to a peak and a steep decrease to near 0 cycles; in bone 24L, the peak in angle is about 0.08 cycles at about 2400 Hz. These features in data for bone 24L occur across all preparations. In each ear, for all TM conditions,

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Voss *et al.*: Non-ossicular signal transmission in human ears   2143

TABLE III. Features of H_{WPD} discussed in the text for each measurement. The columns labeled "Bone" and "Perf. Diam" indicate the preparation number and the state of the TM as intact, removed, or by the diameter in mm of the perforation; the column labeled $\left|\frac{P_{\mathrm{RW}}}{P_{\mathrm{OW}}}\right|$ (max) indicates the magnitude and frequency location of the maximum; the column labeled $\angle\frac{P_{\mathrm{RW}}}{P_{\mathrm{OW}}}$ (max) indicates the angle and frequency location of the maximum in angle; the column labeled $\left|\frac{P_{\mathrm{RW}}}{P_{\mathrm{OW}}}\right|$ (min) indicates the magnitude and frequency location of the minimum; the column labeled $|H_{\mathrm{WPD}}|$ (max) indicates the magnitude and frequency location of the maximum; the column labeled $|Z_{\mathrm{CAV}}|$ indicates the frequencies of the maximum and minimum in $|Z_{\mathrm{CAV}}|$ for each preparation, as reported in Voss *et al.* (2001c). The entry "NA" indicates the measurement was not available.

		Measures of salient features related to H_{WPD}			
Preparation		$\left|\frac{P_{\mathrm{RW}}}{P_{\mathrm{OW}}}\right|$ (max)		$\angle\frac{P_{\mathrm{RW}}}{P_{\mathrm{OW}}}$ (max)		$\left|\frac{P_{\mathrm{RW}}}{P_{\mathrm{OW}}}\right|$ (min)		$|H_{\mathrm{WPD}}|$ (max)		$|Z_{\mathrm{CAV}}|$	
Bone No.	Perf. Diam. (mm)	Mag.	$f_{\max}$ (kHz)	Angle (cycles)	$f_{\angle\max}$ (kHz)	Mag.	$f_{\min}$ (kHz)	Mag.	$f_{\max}$ (kHz)	$f_{Z_{\max}}$ (kHz)	$f_{Z_{\min}}$ (kHz)
	Intact	NA	NA	NA	NA	NA	NA	NA	NA		
	0.1	1.1	0.95	0.02	0.83	1.0	0.78	0.2	0.85	2.08	0.98
9	1.1	1.1	0.85	0.03	0.81	0.9	0.71	0.2	0.81		
	1.9	1.2	0.95	0.05	0.90	0.9	0.85	0.3	0.90		
	3.3	1.1	1.20	0.04	1.12	0.9	1.03	0.2	1.12		
	Intact	1.4	1.29	0.10	1.37	0.8	1.29	0.7	1.42		
	0.5	1.3	1.15	0.06	1.32	0.9	1.15	0.4	1.39	2.67	1.32
13	0.8	1.7	1.27	0.12	1.39	0.8	1.27	0.9	1.46		
	1.4	1.7	1.12	0.14	1.20	0.7	1.12	1.1	1.25		
	1.8	1.9	1.17	0.17	1.27	0.6	1.17	1.3	1.29		
	Removed	1.3	1.32	0.07	1.42	0.8	1.32	0.5	1.44		
	Intact	1.4	2.49	0.09	2.42	0.79	2.29	0.6	2.44		
18	0.4	1.3	2.44	0.06	2.22	0.86	2.05	0.4	2.32	NA	NA
	0.5	1.4	2.37	0.07	2.20	0.84	2.05	0.5	2.27		
	1.0	1.3	2.42	0.07	2.17	0.80	2.03	0.4	2.25		
	Intact	1.2	2.7	0.05	2.08	0.9	1.90	0.28	2.12		
	0.6	1.3	2.4	0.05	1.98	0.9	1.61	0.30	2.37	2.93	1.98
19	1.3	1.2	2.4	0.06	1.90	0.9	1.66	0.42	1.93		
	1.9	1.2	2.4	0.07	1.88	0.9	1.66	0.43	1.93		
	2.6	1.2	2.4	0.07	1.88	0.9	1.64	0.45	1.93		
	Removed	1.2	2.4	0.04	2.05	0.9	1.90	0.25	2.03		
	Intact	1.1	2.34	0.01	2.12	0.9	2.05	0.1	2.08		
	0.7	1.1	2.42	0.02	2.05	0.9	2.00	0.2	2.03	3.66	2.22
	0.7	1.1	3.00	0.02	2.05	0.9	1.95	0.1	2.00		
	1.0	1.1	2.95	0.03	2.61	0.9	1.90	0.2	2.61		
20	1.7	1.1	2.95	0.03	2.56	0.9	1.86	0.2	2.56		
	3.0	1.1	2.95	0.04	2.54	0.9	1.86	0.3	2.56		
	4.0	1.1	2.93	0.04	2.54	0.9	1.83	0.3	2.56		
	5.0	1.1	2.95	0.04	2.54	0.9	2.00	0.3	2.56		
	Removed	1.1	2.95	0.04	2.54	1.0	2.05	0.3	2.54		
	Intact	1.5	2.44	0.08	2.29	0.8	2.10	0.7	2.37		
	0.6	1.4	2.44	0.04	2.27	0.8	2.05	0.4	2.42	3.37	2.20
	1.2	1.7	2.42	0.09	2.25	0.7	2.05	0.8	2.37		
22L	2.0	1.7	2.44	0.09	2.29	0.8	2.03	0.9	2.39		
	3.3	1.7	2.44	0.09	2.32	0.8	2.00	0.9	2.42		
	5.0	1.7	2.44	0.09	2.29	0.8	1.95	0.9	2.39		
	Removed	1.6	2.44	0.07	2.34	0.8	1.93	0.7	2.42		
	Intact	1.2	2.47	0.08	2.37	0.7	2.22	0.5	2.39		
	0.5	1.5	2.44	0.07	2.34	0.8	2.10	0.6	2.42	2.98	2.10
	0.8	1.4	2.49	0.05	2.42	0.8	2.00	0.5	2.49		
	1.2	1.5	2.49	0.07	2.39	0.9	2.03	0.7	2.47		
24L	2.0	1.5	2.49	0.07	2.42	0.8	1.83	0.6	2.47		
	3.3	1.4	2.51	0.07	2.44	0.8	2.00	0.6	2.49		
	5.0	1.6	2.47	0.06	2.37	0.8	1.93	0.8	2.44		
	Removed	1.4	2.49	0.05	2.49	0.8	1.88	0.5	2.49		

Voss *et al.*: Non-ossicular signal transmission in human ears

TABLE III.   (*Continued.*)

		$\left|\dfrac{P_{RW}}{P_{OW}}\right|$ (max)		$\angle\dfrac{P_{RW}}{P_{OW}}$ (max)		$\left|\dfrac{P_{RW}}{P_{OW}}\right|$ (min)		$|H_{WPD}|$ (max)		$|Z_{CAV}|$	
Bone No.	Perf. Diam. (mm)	Mag.	f_{max} (kHz)	Angle (cycles)	$f_{\angle max}$ (kHz)	Mag.	f_{min} (kHz)	Mag.	f_{max} (kHz)	$f_{Z_{max}}$ (kHz)	$f_{Z_{min}}$ (kHz)
	Intact	2.0	2.83	0.23	2.73	0.3	2.69	1.5	2.78		
	0.5	1.9	2.81	0.36	2.71	0.5	2.64	2.2	2.76	4.00	2.61
	0.6	1.7	2.76	0.18	2.64	0.8	1.93	1.4	2.71		
24R	1.3	1.8	2.93	0.37	2.71	0.1	2.76	1.2	2.67		
	2.3	1.8	2.93	0.35	2.69	0.6	2.76	1.6	2.66		
	4.0	1.9	2.93	0.43	2.71	0.1	2.73	1.2	2.64		
	5.0	1.8	2.93	0.34	2.69	0.0	2.71	1.1	2.93		
	Removed	1.5	2.93	0.38	2.69	0.1	2.73	1.1	2.69		

Measures of salient features related to H_{WPD}
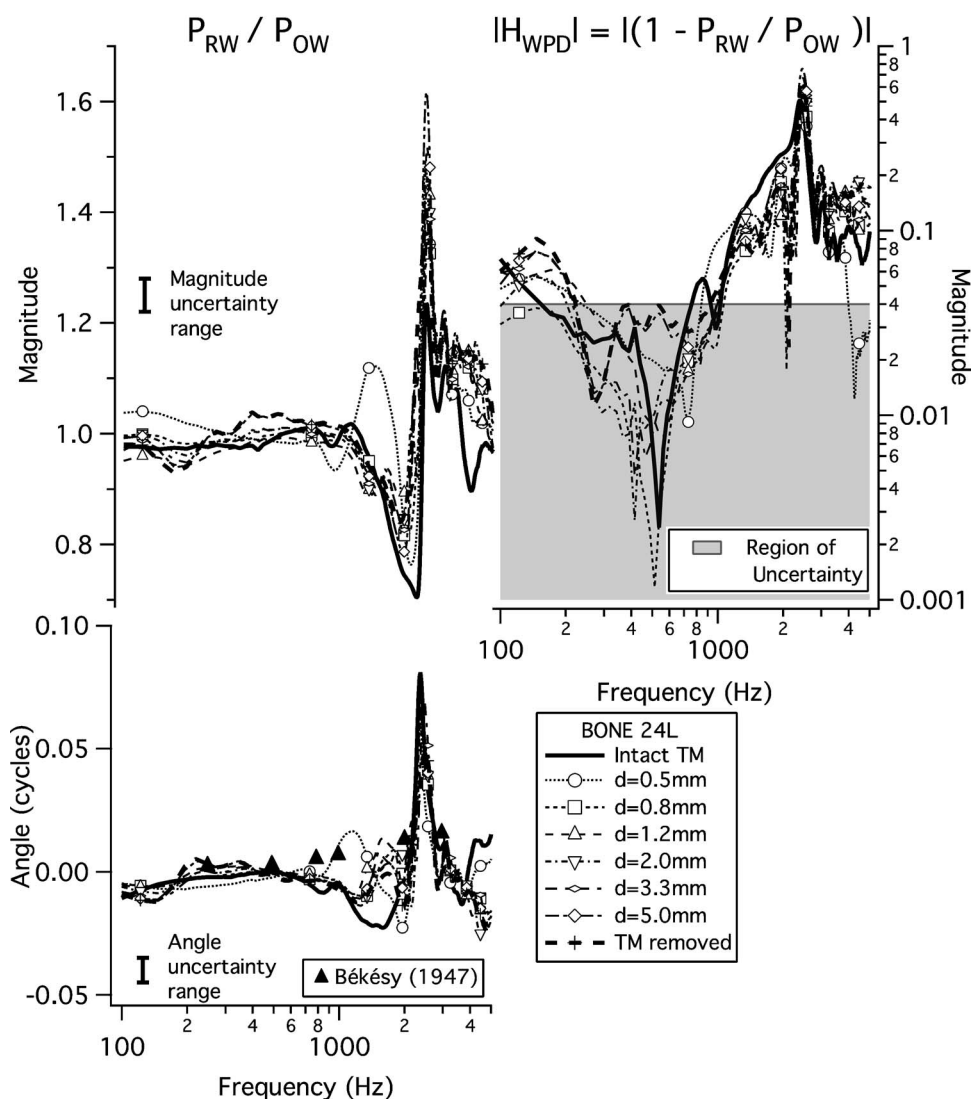
FIG. 4. Example results from bone 24L yielding the window pressure difference transfer ratio H_{WPD}. The parameter is perforation size; symbols are plotted at every 20th data point. Left: P_{RW}/P_{OW}. The vertical bars labeled "Magnitude uncertainty range" and "Angle uncertainty range" indicate the range within which results are not distinguishable with our estimate of measurement error. The magnitude range is 0.06, corresponding to our error estimate of ±0.25 dB, and the angle range is 0.01 cycles, corresponding to our error estimate of ±0.005 cycles. In the lower plot, the filled triangles indicate angle-difference measurements from Békésy (1947) between the oval- and round-window pressures at six frequencies in one temporal bone with the TM removed. Right: $|H_{WPD}|=|1-P_{RW}/P_{OW}|$ calculated from data on left. Points below 0.04 (gray shaded region) may have sizable errors (see Sec. III C 4).

between about 1000 and 3000 Hz, the ratio exhibits a local magnitude minimum adjacent to a local magnitude maximum, with an increase and decrease in angle. Although this pattern is seen in all ears, there are inter-ear variations in size and frequency of peaks (Table III).

Although at most frequencies P_{RW} and P_{OW} are roughly equal, between 1000 and 3000 Hz $|P_{RW}|$ and $|P_{OW}|$ have large derivatives with respect to frequency, and the biggest differences between P_{RW} and P_{OW} occur at the frequencies where $|P_{RW}|$ and $|P_{OW}|$ change rapidly (i.e., near 2500 Hz). All measurements show that in the frequency region of steep change a decrease in $|P_{RW}|$ occurs in a slightly lower frequency range than the decrease in $|P_{OW}|$. This "frequency shift" produces the pattern of the local minimum in the pressure ratio $|P_{RW}/P_{OW}|$ at a frequency just below a local maximum in $|P_{RW}/P_{OW}|$. The consistency of this behavior suggests that the dominant mechanism(s) is the same in all preparations. Possibilities for this mechanism(s) are proposed in Sec. V C.

In Sec. III C 4 we suggest that errors in the measurements of the oval- and round-window pressures lead to uncertainties of ±0.25 dB in magnitude and ±0.005 cycles in angle for the calculation of P_{RW}/P_{OW}. The corresponding "uncertainty ranges" of ±0.03 in magnitude (assuming $P_{RW} \approx P_{OW}$) and ±0.005 cycles in angle are indicated by a bracket in each plot of Fig. 4 (left). At a given frequency, measurements that are within the span of this bar may not represent a real difference. Therefore, differences between perforation conditions that are smaller than a factor of 0.06 in magnitude and 0.01 cycles in angle are considered not detectably different. Most of the measurements of magnitudes and angles of P_{RW}/P_{OW} are within the uncertainty estimate, so the small differences with perforation size may not reflect real differences. Measurements outside this uncertainty estimate, such as those from bone 24L between 1000 and 3000 Hz for all perforation sizes, are presumed to be significant.

Next, we consider the window-pressure-difference transfer function H_{WPD} (Fig. 4, right). This transfer function is simply related to the pressure ratio: $H_{WPD} = 1 - (P_{RW}/P_{OW})$. Thus, when $P_{RW}/P_{OW} \approx 1$, the magnitude $|H_{WPD}| \ll 1$ and the angle $\angle H_{WPD}$ can be very sensitive to small changes in magnitude.

Most of the $|H_{WPD}|$ results increase with frequency above 800 Hz to a local maximum between 1000 and 3000 Hz. This local maximum is a consequence of the local maximum of the pressure ratio $|P_{RW}/P_{OW}|$. However, below 800 Hz, $0.97 \leq |P_{RW}/P_{OW}| \leq 1.03$, such that $|H_{WPD}| < 0.04$, and we cannot reliably determine $|H_{WPD}|$.

2. Effects of perforation size on $|H_{WPD}|$

For a given frequency below 1000 Hz, variations in bone 24L (Fig. 4) and the other seven preparations (Voss, 1998, page 251) do not seem to be systematically related to perforation size. The variability may result from small errors in our pressure measurements when $P_{OW} \approx P_{RW}$, as the estimate of the lower limit for which we can accurately measure $|H_{WPD}|$ is 0.04 (Sec. III C 4), which is indicated on the plot in Fig. 4 (right). Our calculations of $|H_{WPD}|$ below about

1000 Hz are in the "region of uncertainty." Data in this region are shown because their coherence across measurements in this region indicates that the main trends may be representative even at levels below $|H_{WPD}| = 0.04$.

To compare results across ears and perforation size we systematically sort P_{RW}/P_{OW} and H_{WPD} results into seven groups differing in TM condition (defined in the legend of Fig. 2), and we compute the means within each of the seven groups for the three quantities $|P_{RW}/P_{OW}|$, $\angle(P_{RW}/P_{OW})$ and $|H_{WPD}|$. To determine if the differences between these means are significant, at each of six frequencies (125, 250, 500, 1000, 2000, and 4000 Hz) we use the "resampling-permutation" methodology described in Sec. III D to compute p values for differences between each group's mean relative to all other measurements.

Tests between all combinations of seven TM conditions at six frequencies and for three quantities yield a total of 378 p values, which are all reported in Voss (1998). Almost all of the computed p values for all three quantities $(|P_{RW}/P_{OW}|, \angle(P_{RW}/P_{OW}), |H_{WPD}|)$ are such that $p \gg 0.05$, indicating that the differences between the means are not significant. There are comparisons for which $p < 0.05$, but these do not suggest any trends with perforation size. For example, there are four comparisons (of a total of 126 comparisons) for $|P_{RW}/P_{OW}|$ for which $p < 0.05$; two of these occur at 125 Hz and the other two at 250 Hz, both frequencies where the uncertainties limit the measurements. Similarly, occasional values where $p < 0.05$ occur in the calculations for $\angle(P_{RW}/P_{OW})$ at the lower frequencies. However, there are no cases where $p < 0.05$ for $|H_{WPD}|$. Thus, in general, we conclude that our results for $|P_{RW}/P_{OW}|$, $\angle(P_{RW}/P_{OW})$, and $|H_{WPD}|$ do not demonstrate dependence on perforation size for any frequency range.

3. Effects of perforation location on H_{WPD}

We compare the effects of perforation location (i.e., anterior-inferior versus posterior-inferior) on the quantities $|P_{RW}/P_{OW}|$, $\angle(P_{RW}/P_{OW})$, and $|H_{WPD}|$ in a similar manner to our comparison of perforation size (Sec. IV C 2). First, we form three perforation-size categories: (1) $0.1 \leq d \leq 1$ mm; (2) $1 < d \leq 2$ mm; and (3) $2 < d \leq 4$ mm. Next, within each size category, means are computed separately (at six frequencies) for $|P_{RW}/P_{OW}|$, $\angle(P_{RW}/P_{OW})$, and $|H_{WPD}|$ from perforations in the anterior-inferior quadrant and posterior-inferior quadrant. These means are then tested to determine if the differences between the two locations are significant.

The resulting p values for the perforation-location dependence of $|P_{RW}/P_{OW}|$, $\angle(P_{RW}/P_{OW})$, and $|H_{WPD}|$ do not show systematic dependences of H_{WPD} on perforation location. For 51 of the 54, $p > 0.05$; the remaining three comparisons, where $p < 0.05$, do not suggest any trends. For example, for the diameter range $1 < d \leq 2$ mm, there is a possible location dependence for $\angle(P_{RW}/P_{OW})$ between anterior and posterior perforations for the frequencies 500 and 1000 Hz. However, this dependence is not seen for smaller or larger perforations at these frequencies. In general, the results do not indicate a perforation-location dependence of any of the three quantities $|P_{RW}/P_{OW}|$, $\angle(P_{RW}/P_{OW})$, and $|H_{WPD}|$.

Voss *et al.*: Non-ossicular signal transmission in human ears
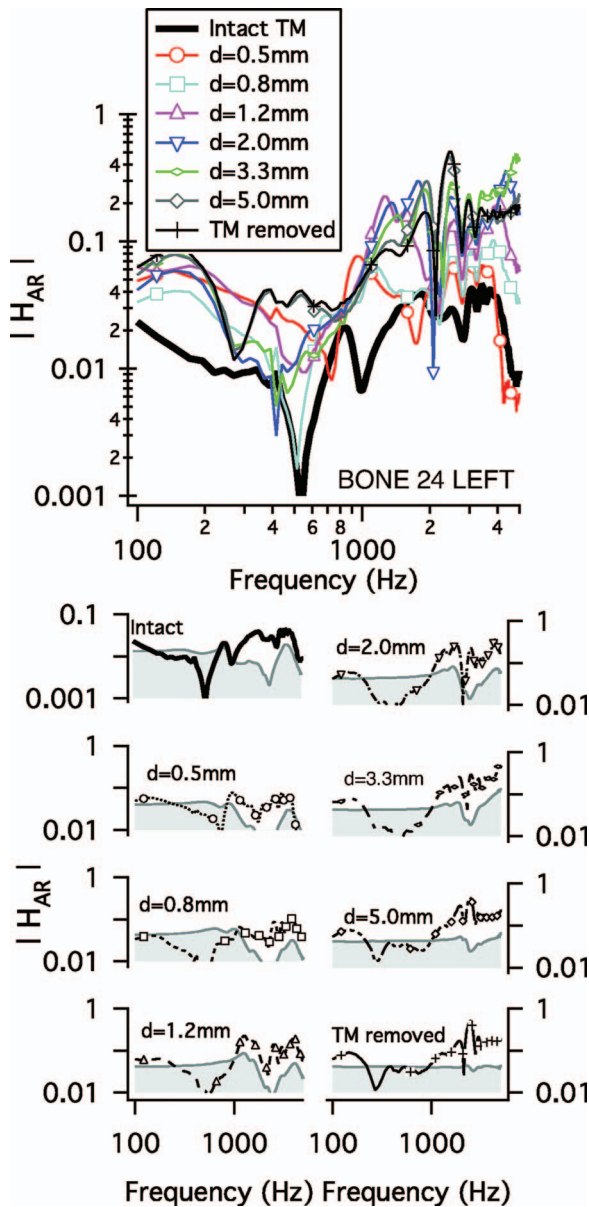
FIG. 5. Acoustic-route transmission magnitude $|H_{AR}|$. Upper: $|H_{AR}|$ for bone 24L. The parameter is perforation size as indicated in the legend. Symbols indicate every 20th data point. Lower: $|H_{AR}|$ for each TM condition with our estimate of measurement "region of uncertainty," which is shaded in gray. For each case, the upper limit of this region is $0.04|H_{P_{CAV}}|$ (Table I) where the $|H_{P_{CAV}}|$ (Fig. 2) corresponds to the relevant TM condition. Note that the displayed y-axis range differs between the intact and perforated TM cases to accommodate differences in $|H_{P_{CAV}}|$.

D. Effects of tympanic-membrane perforations on net sound transmission via the acoustic route: $|H_{AR}|$

1. Measurements of $|H_{AR}|$

The acoustic-route transfer function $|H_{AR}|$ is plotted in Fig. 5 for the example bone (24L) [and in Voss (1998) for all eight bones]. The lower plots in Fig. 5 (corresponding to the TM intact, removed, and with six perforation sizes) show that, below 1000 Hz, the measurements of $|H_{AR}|$ are often in the "region of uncertainty," which is based on relative errors in the measurement of P_{OW} and P_{RW}. Thus, some of these low-frequency measurements of $|H_{AR}|$ are likely to be influenced by measurement errors.
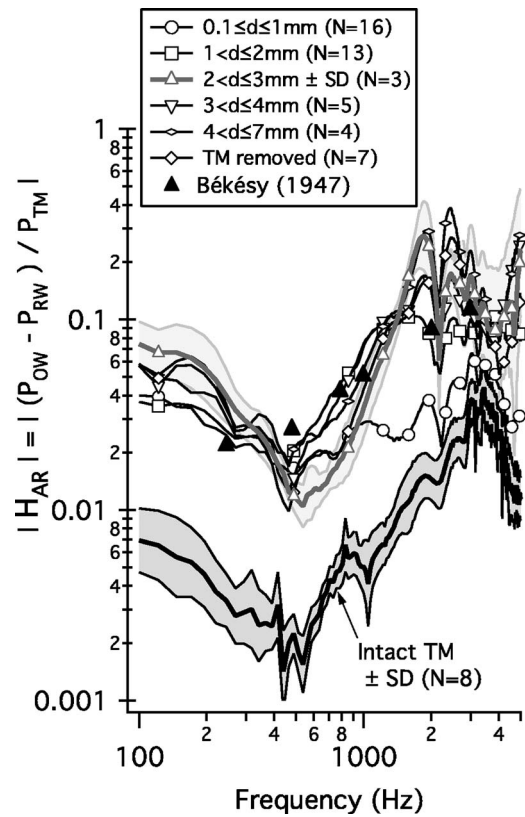
FIG. 6. Mean acoustic-route transfer function magnitude $|H_{AR}|$ computed in the logarithmic domain for the indicated TM conditions. The standard errors for the intact TM and perforation-diameter range $2 < d \leq 3$ mm are indicated by the shaded regions. The standard errors for all other groups are about the same or smaller than those for $2 < d \leq 3$ mm. Symbols indicate every 20th data point. The six black triangles indicate measurements from Békésy (1947) of the window pressure difference from one temporal bone with the TM removed. Below 1000 Hz, values may be influenced by relative errors in the measurement of the oval- and round-window pressures; for these lower frequencies the plotted curves can be considered an upper bound on $|H_{AR}|$.

2. Effect of perforation size on $|H_{AR}|$

Figure 5 shows two general trends that occur across the preparations with perforation size as a parameter. First, with an intact TM $|H_{AR}|$ is generally less than $|H_{AR}|$ with a perforation; the increase in $|H_{AR}|$ with a perforation is a consequence of the perforation-induced increase in $|H_{P_{CAV}}|$ as seen in Fig. 2. Second, for frequencies above about 1000 Hz, the perforation size influences $|H_{AR}|$ in a manner similar to $|H_{P_{CAV}}|$ (see Fig. 2): at most frequencies, $|H_{AR}|$ increases with perforation size, with the largest increase associated with the change from less than 1 mm to more than 1 mm in perforation diameter. These trends are apparent in the means of all $|H_{AR}|$ measurements, grouped by perforation size (Fig. 6).

We performed a probabilistic analysis to determine whether the differences between the means of the seven different groups plotted in Fig. 6 are significant. Using the "resampling-permutation" methodology (Sec. III D), we computed null-hypothesis probabilities p for each $|H_{AR}|$ mean measurement relative to the six other mean measurements. The resulting probabilities that test for significance of differences in the means of $|H_{AR}|$ are summarized here. Gen-

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Voss et al.: Non-ossicular signal transmission in human ears   2147
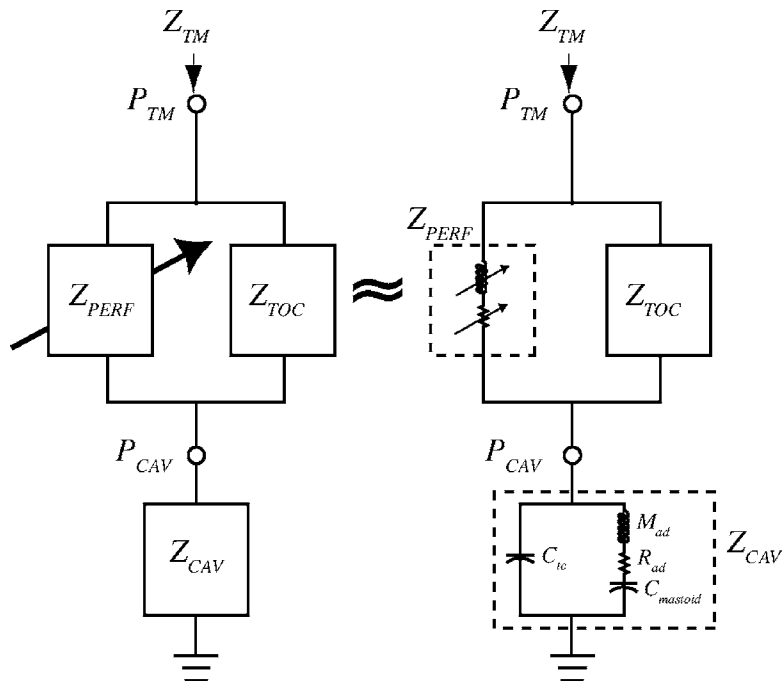
FIG. 7. Analog electric-circuit model of the human middle ear with a TM perforation. Voltages are analogous to sound pressures. The impedance of the TM loaded by the ossicular chain is represented by Z_{TOC}. The perforation is represented as the impedance Z_{PERF}; the impedance of the middle-ear cavities is represented as Z_{CAV}; P_{TM} is the sound pressure at the TM; P_{CAV} is the sound pressure in the middle-ear cavity; Z_{TM} is the input impedance in the ear canal at the TM. Arrows on circuit components indicate changes with perforation size. Left: General topology [as developed in Voss *et al.* (2001c)]. Right: Representation by ideal elements that represent the impedances Z_{PERF} [see Voss *et al.* (2001c) for detailed description] and Z_{CAV} [see Voss *et al.* (2000) and Stepp and Voss (2005) for detailed description] C_{tc} represents the acoustic compliance of the tympanic cavity; M_{ad} represents the acoustic mass of the aditus; R_{ad} represents the resistance of the aditus; $C_{mastoid}$ represents the compliance of the antrum and mastoid space.

erally, the results are consistent with the qualitative impression given by Fig. 6. In particular, the differences between $|H_{AR}|$ with an intact TM and $|H_{AR}|$ with a TM perforation of any size are highly significant ($p \ll 0.05$) for nearly all perforation sizes at nearly all tested frequencies. At frequencies below 1000 Hz, there are no significant differences ($p < 0.05$) between perforations of different sizes; either there are no true differences or our analysis is limited by inaccuracy in measuring the small pressure difference at these lower frequencies. Above 1000 Hz, there is evidence of significant differences in $|H_{AR}|$ between the larger and smaller perforations: for example, at 2000 Hz there are significant differences between 15 of the 16 comparisons for size categories up to 3 mm diameters. There are not systematic trends at the other tested frequencies (1000 and 4000 Hz), and comparisons at 2000 Hz are not significant for comparisons among size pairs for which both are greater than 3 mm in diameter.

3. Effect of perforation location on $|H_{AR}|$

We compare the effects of perforation location on the $|H_{AR}|$ in a similar manner to our comparison of perforation size performed above for P_{RW}/P_{OW} and H_{WPD} (Section IV C 2). Again, we compare perforations in the anterior-inferior quadrant to those in the posterior-inferior quadrant. First, we form three perforation-size categories: (1) $0.1 \leq d \leq 1$ mm; (2) $1 < d \leq 2$ mm; and (3) $2 < d \leq 4$ mm. All perforations such that $0.1 \leq d \leq 4$ mm are placed into the appropriate category. Next, within each size category, means are computed separately for $|H_{AR}|$ from perforations in the

anterior-inferior quadrant and posterior-inferior quadrant. These means are then tested to determine if their differences are significant. There is no evidence of any perforation-location dependence for $|H_{AR}|$, as for all tests $p > 0.20$.

V. DISCUSSION

A. Summary of measurements and their limitations

We have represented the acoustic-route transfer function $H_{AR} = (P_{OW} - P_{RW})/P_{TM}$ as the product of two transfer functions: the cavity-pressure transfer function ($H_{P_{CAV}}$) and the window-pressure difference transfer function (H_{WPD}); i.e., $H_{AR} = H_{P_{CAV}} H_{WPD}$. Here we summarize our conclusions regarding the effects of perforation location and size on these transfer functions.

The measurements show that $H_{P_{CAV}}$ depends on perforation size and not perforation location. In contrast, measurements of H_{WPD} do not show dependence on (1) whether the TM is intact or perforated, (2) perforation size, or (3) perforation location. Below 1000 Hz, the measurement accuracy for H_{WPD} and H_{AR} may be degraded by errors in the measurement of $P_{OW} - P_{RW}$ (Sec. III C 4), but the measurements provide an upper limit for $|H_{WPD}|$.

B. Acoustic mechanisms that determine $H_{P_{CAV}}$

An analog circuit model representing the acoustic interactions in middle ears with a TM perforation is shown on the left side of Fig. 7. The topology of this model requires that the volume velocity into the cavity is the sum of the volume velocities of the TM and of air through the perforation. Z_{PERF}
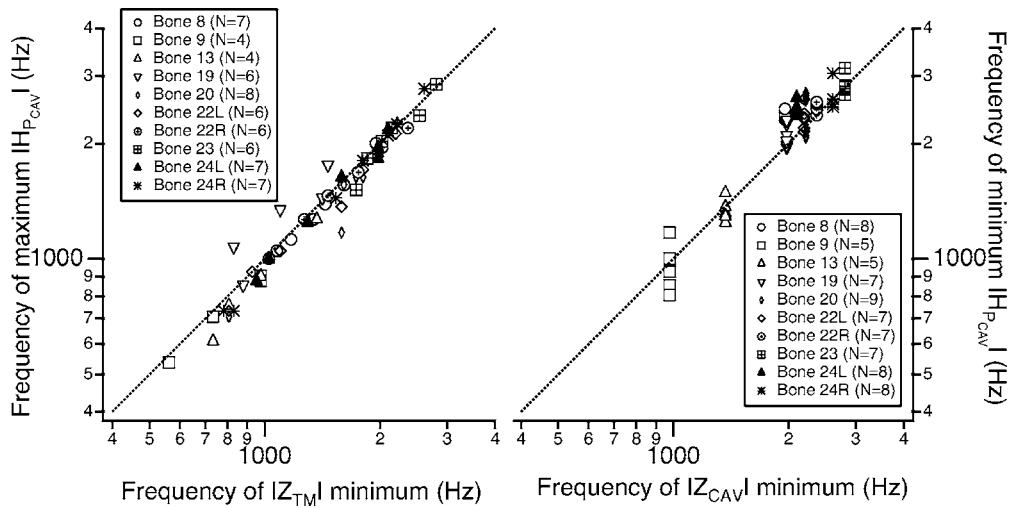
FIG. 8. Scatter plots that connect frequencies of extrema of $|H_{P_\text{CAV}}|$ and frequencies of extrema in the magnitudes of the middle-ear input impedance $|Z_\text{TM}|$ and the middle-ear cavity impedance $|Z_\text{CAV}|$. (Impedance measurements were reported in Voss *et al.* (2001c).) In each plot points are shown for all measurements including several perforation sizes (see legends for numbers) and all preparations. Bone 18 is omitted because impedance measurements do not exist for that preparation. The straight lines in each plot represent the relationship y (ordinate)=x (abscissa). Left: the frequency of the maximum in $|H_{P_\text{CAV}}|$ versus the frequency of the minimum of $|Z_\text{TM}|$. Right: The frequency of the minimum in $|H_{P_\text{CAV}}|$ versus the frequency of the minimum of the impedance of the middle-ear cavity $|Z_\text{CAV}|$. N.B. The horizontal coordinate for all points in a given ear is the same, because Z_CAV was measured in only one condition—i.e., with the TM removed.

and Z_CAV are represented on the right side of Fig. 7 by simple connections of ideal elements based on structural features of the perforation and cavities, and acoustic theory, as detailed in other work (Voss *et al.*, 2000, 2001c; Stepp and Voss, 2005). This model topology is used here to consider mechanisms that might influence H_{P_CAV}.

With this model (Fig. 7, left) we can relate impedance and pressure quantities that we have measured, specifically the circuit configuration requires that

$$H_{P_\text{CAV}} \equiv \frac{P_\text{CAV}}{P_\text{TM}} = \frac{Z_\text{CAV}}{Z_\text{TM}}, \tag{7}$$

where Z_CAV is the impedance of the middle-ear cavities and Z_TM is the input impedance at the tympanic membrane.

Equation (7) suggests that, if $|Z_\text{CAV}|$ varies gradually with frequency, then a local maximum in $|H_{P_\text{CAV}}|$ might occur at a local minimum in $|Z_\text{TM}|$. This relationship is tested in Fig. 8 (left) with independent measurements of $|H_{P_\text{CAV}}|$ (reported here) and $|Z_\text{TM}|$ [reported in Voss *et al.* (2001c)]. The scatter plot demonstrates that the frequency of the local minimum in the magnitude of the input impedance at the TM Z_TM correlates well with the frequency of the local maximum in $|H_{P_\text{CAV}}|$. Note that variations in each ear are a result of changes in perforation size, which alter both coordinates. Thus, the low-frequency (0.5–2.8 kHz) maximum in $|H_{P_\text{CAV}}|$ depends on features of both the cavities and the perforation (as Z_TM depends on both quantities).

Similarly, the model suggests that the frequency of the minimum in the cavities' impedance magnitude $|Z_\text{CAV}|$ [reported in Voss *et al.* (2000) and Voss *et al.* (2001c)] might occur at the same frequency as the minimum in $|H_{P_\text{CAV}}|$. This relationship is tested in Fig. 8 (right) with independent measurements of $|H_{P_\text{CAV}}|$ and $|Z_\text{CAV}|$ [reported in Voss *et al.* (2000)]. The scatter plot demonstrates that the frequency of the local minimum in the cavities' impedance magnitude $|Z_\text{CAV}|$ correlates well with the frequency of the local mini-

mum in $|H_{P_\text{CAV}}|$. Thus, the minimum in $|H_{P_\text{CAV}}|$ (0.9–2.7 kHz) depends primarily on features of the middle-ear cavities, consistent with the observation that the minimum frequency appears to be independent of perforation diameter.

The more detailed model in Fig. 7 (right) provides representations for the model impedances Z_PERF and Z_CAV (Fig. 7, left) (Voss *et al.*, 2000, 2001c). The perforation is represented by a combination of mass and resistive components that depend on frequency and perforation size (Voss *et al.*, 2001c); except for the very smallest perforations and the lowest frequencies (diameters less than 1 mm and frequencies less than 300 Hz), the perforation's impedance is mass dominated such that the impedance magnitude $|Z_\text{PERF}|$ generally can be represented as

$$Z_\text{PERF} \approx j\omega M_\text{PERF}, \tag{8}$$

where $\omega = 2\pi f$ with f frequency and M_PERF the equivalent acoustic mass of the perforation. We note that $|Z_\text{PERF}|$ increases with frequency for a constant diameter and decreases with larger perforation diameters for a constant frequency [see Fig. 3 of Voss *et al.* (2001c)]. The model representation for the middle-ear cavities includes two volume-velocity paths: (1) the tympanic-cavity air space, represented by a compliance, and (2) the aditus-ad-antrum leading to the antrum and mastoid-cell network represented by a series mass and resistance leading to a compliance. [This model representation is detailed in Voss *et al.* (2000) and Stepp and Voss (2005)]. For low frequencies[3]

$$Z_\text{CAV} \approx \frac{1}{j\omega C_\text{CAV}}, \tag{9}$$

where C_CAV is a compliance that is proportional to the cavities' equivalent volume. The mass of the perforation and the compliance of the middle-ear cavities produce a series reso-

nance (i.e., $Z_{PERF} + Z_{CAV} = 0$) such that $|H_{P_{CAV}}|$ is a maximum at

$$f_{max} = \frac{1}{2\pi C_{CAV} M_{PERF}}, \quad (10)$$

where f_{max} increases with increasing perforation diameter because M_{PERF} decreases with increasing perforation diameter. This dependence is qualitatively consistent with the observation that the frequency of the maximum in $|H_{P_{CAV}}|$ increases with increasing perforation diameter (Table II). Thus, the model predicts that the frequency of the maximum in $|H_{P_{CAV}}|$ is effectively determined by the dimensions of the perforation and the volume of the middle-ear air space.

For each ear, the frequency of the minimum in $|H_{P_{CAV}}|$ changes moderately with perforation size and is correlated with a feature of Z_{CAV} (Fig. 8, right). The model for Z_{CAV} (Fig. 7, right) has a minimum in Z_{CAV} that occurs at the frequency of the series resonance between the mass of the aditus (M_{ad}) and the compliance of the air in the antrum and mastoid ($C_{mastoid}$) (Voss *et al.*, 2000; Stepp and Voss, 2005) such that the frequency of the minimum in $|Z_{CAV}|$ occurs at

$$f_{min} = \frac{1}{2\pi C_{mastoid} M_{ad}}. \quad (11)$$

Thus, the model predicts that the frequency of the minimum in $|H_{P_{CAV}}|$ is effectively determined by the resonance between the mass that represents the aditus ad antrum and the compliance that is determined by the volume of the antrum and mastoid air space. This dependence is consistent with the observation that the frequency of the minimum in $|H_{P_{CAV}}|$ does not depend on perforation size (Table II). We note that there is substantial variation in f_{min} among preparations, which appears to result from intra-ear structural variations in the aditus, antrum, and mastoid.

Our $H_{P_{CAV}}$ results can be compared to finite-element (FE) model predictions from Gan *et al.* (2006). The FE model pressures that determine $H_{P_{CAV}}$ are plotted in Fig. 4 of Gan *et al.* (2006) for an intact TM and in Figs. 6–8 of Gan *et al.* (2006) for small TM perforations; the comments here focus on the model predictions corresponding to pressures 2 mm from the umbo and within the middle-ear cavity. The FE model results are consistent with our measurements in that the model pressures at both the oval and round windows are nearly identical, making the calculation of $H_{P_{CAV}}$ essentially independent of the location of measurement of middle-ear pressure. For an intact TM, FE model results of Gan *et al.* (2006) differ from both our measurements and our lumped-element model at frequencies below 1000 Hz. Both the lumped-element model of Fig. 7 and measurements of $H_{P_{CAV}}$ on 11 ears with intact TMs [Fig. 2 here and Appendix G of Voss (1998)] show that the magnitude $|H_{P_{CAV}}|$ is relatively independent of frequency for low frequencies (typically below 500–1000 Hz) with a corresponding angle $\angle H_{P_{CAV}}$ that is nearly zero; Fig. 4 of Gan *et al.* (2006) shows low-frequency FE model predictions that result in $|H_{P_{CAV}}|$ systematically increasing with frequency and $\angle H_{P_{CAV}}$ systematically decreasing with frequency. For a perforated TM, the FE model results [Figs. 6–8 of Gan *et al.* (2006)] are
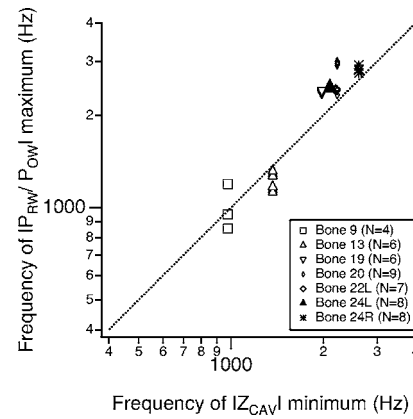
FIG. 9. Scatter plot that connects the window-pressure ratio to features of the middle-ear cavity. Plotted is the frequency of the magnitude maximum in $|P_{RW}/P_{OW}|$ versus the frequency of the minimum of the middle-ear cavity impedance $|Z_{CAV}|$. (Impedance measurements were reported in Voss *et al.* (2001c).) In many cases points from different perforations on the same preparation (bone) are identical and so it is not possible to distinguish all of the plotted points. Data are included only from ears in which variability in P_{RW}/P_{OW} calibrations were small; Bone 18 does not appear as impedance measurements are not available.

similar to both our measurements and our lumped-element model at all frequencies; specifically, magnitude $|H_{P_{CAV}}|$ is approximately 1 for low frequencies (typically below 1000 Hz) with a corresponding angle $\angle H_{P_{CAV}}$ that is nearly zero, consistent with the mechanisms described in detail by Voss *et al.* (2001c) for how a perforation affects the pressure difference across the TM.

C. Acoustic mechanisms that determine H_{WPD}

The quantities P_{OW}, P_{RW}, and H_{WPD} are not represented by the lumped-element model of Fig. 7, as the model assumes a uniform middle-ear-cavity pressure. However, we can suggest a mechanism that relates $|P_{RW}/P_{OW}|$ to the middle-ear cavities. The results show a consistent peak in $|P_{RW}/P_{OW}|$ at a frequency that varies among preparations between 800 and 3000 Hz but in a given preparation is independent of perforation size. The lack of dependence on perforation size suggests that this peak is related to the structure of the middle-ear cavities. The frequency of the peak in $|P_{RW}/P_{OW}|$ and the frequency of the minimum in $|Z_{CAV}|$ are closely correlated (Fig. 9). At the frequency of the minimum in $|Z_{CAV}|$, one might expect a maximum in volume velocity (and particle velocity) entering the aditus from the tympanic cavity. This velocity maximum might produce a relatively large spatial derivative in pressure[4] in the region of the windows and therefore an increased difference between P_{RW} and P_{OW}. Thus, a feature of the middle-ear cavity structure might determine the peak frequency of $|P_{RW}/P_{OW}|$.

In the future, finite-element models might provide additional insight into the relationship between P_{OW} and P_{RW}. The plots and text in Gan *et al.* (2006) are consistent with small differences between P_{OW} and P_{RW}.

D. Effects of reduced middle-ear air space

Measurements in Stepp and Voss (2005) demonstrate that reduction of the mastoid portion of the cadaver ears

Voss *et al.*: Non-ossicular signal transmission in human ears

influences middle-ear acoustics. The measurements presented here were made on bones that were removed with a circular saw (Schuknecht, 1968) such that much of the mastoid air-cell network was removed. The demonstrable effects of this loss of mastoid air cells are a decrease in total middle-ear air-space volume and removal of highly individualistic resonances in the cavity impedance at frequencies above 1000 Hz (Stepp and Voss, 2005). These results suggest that, if the acoustic-route measurements reported here had been made on ears with intact mastoid spaces, changes would be:

1. Greater inter-ear variability in $H_{P_{\text{CAV}}}$ as a result of inter-specimen structural variations of mastoid air spaces; specifically, the frequencies and values of the extrema in $H_{P_{\text{CAV}}}$ at higher frequencies (i.e., $f > 1000$ Hz) would be altered.
2. Larger changes between the intact and the perforated $|H_{P_{\text{CAV}}}|$, especially at the lower frequencies ($f < 1000$ Hz) where the acoustics are compliance dominated. Specifically, the larger cavity volumes would lead to smaller low frequency $|H_{P_{\text{CAV}}}|$ with an intact TM, whereas a perforation would still make $H_{P_{\text{CAV}}} \approx 1$ for the lower frequencies.
3. As H_{WPD} is apparently determined by the acoustic configuration of the middle-ear cavities, it seems likely that H_{WPD} would also have different features with intact cavities.

E. Conditions in which the acoustic route may be important for hearing with tympanic-membrane perforations

1. Overview

As discussed in the Introduction, when an ear is normal, the ossicular route is the dominant mechanism for sound transmission to the cochlea, and transmission via the acoustic route is negligibly small. Figure 6 demonstrates that sound transmission via the acoustic route increases when the TM is perforated. Here, we look for situations in ears with perforations in which the acoustic route becomes significant.

Voss et al. (2001b) present measurements of the middle-ear transfer function V_S/P_{TM} (stapes velocity per pressure at the TM) made with both intact and perforated TMs. These measures are assumed to include transmission via both the ossicular and acoustic routes:

$$\frac{V_S^{\text{total}}}{P_{\text{TM}}} = \frac{V_S^{\text{acoustic}}}{P_{\text{TM}}} + \frac{V_S^{\text{ossicular}}}{P_{\text{TM}}}. \qquad (12)$$

In this section, we use our measurements of the acoustic-route transfer function H_{AR} to compute an upper bound for the acoustic-route transmission $|V_S^{\text{acoustic}}/P_{\text{TM}}|$. We then compare this upper bound with our measurements of $|V_S^{\text{total}}/P_{\text{TM}}|$ (Voss et al., 2001b) in order to determine the importance of the acoustic-route component in the total transmission.

2. Comparison of acoustic route transmission to normal middle-ear transmission

In this section, we compare the total middle-ear transmission to the acoustic-route transmission for ears with either intact or missing TMs.

The stapes velocity per pressure at the TM that results from the pressure difference at the oval and round windows, $V_S^{\text{acoustic}}/P_{\text{TM}}$, can be expressed in terms of three quantities: the pressure measurements at the oval and round windows, the stapes-cochlea acoustic impedance Z_{SC}, and the area of the stapes footplate A_S

$$\frac{V_S^{\text{acoustic}}}{P_{\text{TM}}} = \frac{P_{\text{OW}} - P_{\text{RW}}}{P_{\text{TM}} Z_{\text{SC}} A_S} = \frac{H_{\text{AR}}}{Z_{\text{SC}} A_S}. \qquad (13)$$

Here we use the average Z_{SC} measured in cadaveric temporal bones by Merchant et al. (1996) and the average A_S $=3.2$ mm^2 reported by Wever and Lawrence (1954, p. 417).

Equation (13) was used to compute the mean $|V_S^{\text{acoustic}}|/P_{\text{TM}}$ for both the TM intact and TM removed cases (two lower curves in the top plot of Fig. 10); these curves are upper bounds for middle-ear sound transmission with either the TM intact or removed. The magnitude is larger with TM removed because $|H_{\text{AR}}|$ is larger (see Fig. 5). Higher in the top plot (Fig. 10) is the mean measured total middle-ear sound transmission $|V_S^{\text{total}}/P_{\text{TM}}|$ from Fig. 4 of Voss et al. (2000). The magnitude of the total transmission relative to the acoustic-route transmission (in the lower plot of Fig. 10) describes the importance of the acoustic route in sound transmission to the cochlea. With an intact TM, the total stapes velocity is 100–3000 times (i.e., 40–70 dB) greater than the upper bound on the acoustic-route stapes velocity (Fig. 10, lower). Thus, when the TM and ossicular chain are normal, transmission via the acoustic route is negligibly small relative to the ossicular route.

The measurements of H_{AR} allow us to calculate an upper bound for hearing loss with the TM removed; measurements of stapes velocity under this condition are not available because the stapes motion was below the noise floor of the measurement system. H_{AR} for the TM-removed condition is similar to that with TM perforations (Fig. 6). When the TM is removed (or perforated) the upper bound on transmission via the acoustic route $V_S^{\text{acoustic}}/P_{\text{TM}}$ increases (Fig. 10) and the total transmission $V_S^{\text{total}}/P_{\text{TM}}$ decreases (not shown), with specific decreases dependent on the perforation size (Voss et al. 2001b). When the decrease in $V_S^{\text{total}}/P_{\text{TM}}$ is substantial (e.g., with larger perforations), the acoustic-route transmission could make significant contributions to the total transmission and thus determine the hearing loss. This upper hearing-loss limit is equivalent to the ratio between total sound transmission in the intact condition ($V_S^{\text{total}}/P_{\text{TM}}$ with an intact TM) and the upper bound on the acoustic-route sound transmission in the TM removed condition ($V_S^{\text{acoustic}}/P_{\text{TM}}$ with the TM removed). This estimate of an upper limit for loss with TM perforations is included in the lower part of Fig. 10. The limit is between 40 and 60 dB at frequencies up to 2000 Hz, and from 2000 to 4000 Hz between 30 and 40 dB. These limits are comparable to the 45–65 dB hearing losses that Békésy (1936) reported in patients with missing TMs, mallei, and incudes (Békésy, 1936; Peake et al., 1992).

Figure 11 compares total transmission loss with different sized TM perforations (Voss et al., 2001a, b) to an estimated upper limit for transmission loss that is determined by the acoustic route. At most frequencies, even with perforations, the measured stapes-velocity (V_S) losses are above the limit
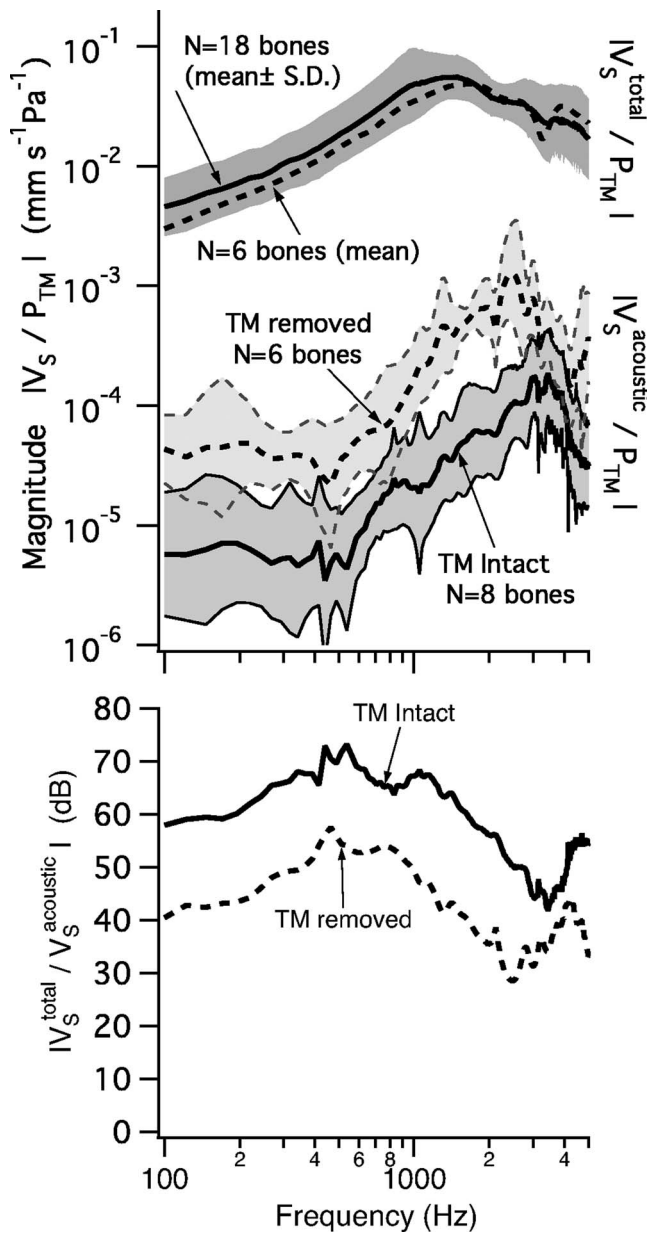
FIG. 10. Results relevant to the possible impact of acoustic coupling. Upper: Eq. (13) was used to compute the mean $|V_S^{acoustic}|/P_{TM}$ for both the TM intact and TM removed cases (two lower curves in the top plot). For comparison the upper curve shows the average measured total middle-ear transmission $|V_S^{total}/P_{TM}|$. $|V_S^{total}/P_{TM}|$ was measured with the laser-Doppler system on bones with an intact TM. The mean±the standard deviation (S.D.), associated with 18 bones, comes from Voss *et al.* (2000, Fig. 4). The mean of $N=6$ normal bones (dashed line upper) is from the six bones used to calculate the acoustic-route stapes velocity with the TM removed (dashed line lower), where $|V_S^{acoustic}/P_{TM}|$ is calculated from Eq. (13) using the mean H_{AR} for the normal and the TM removed conditions. (The closeness of these two upper curves indicates that the smaller group was similar to the larger group in making comparisons with the $N=6$ TM removed group.) The standard deviations are shaded and account for variation in both (1) the H_{AR} measurements made here and (2) the Z_{SC} measurements made by Merchant *et al.* (1996). Lower: Curves show the dB differences between each of the two acoustic-route transmissions and the total transmission curve (from the top plot). The vertical coordinate here represents the hearing loss resulting from loss of ossicular coupling for either TM condition.

set by the acoustic route, indicating that the ossicular route dominates the stapes-velocity measurements. However, the plotted upper limit for the "acoustic-route" suggests that the acoustic route may limit the loss at the lowest frequencies,
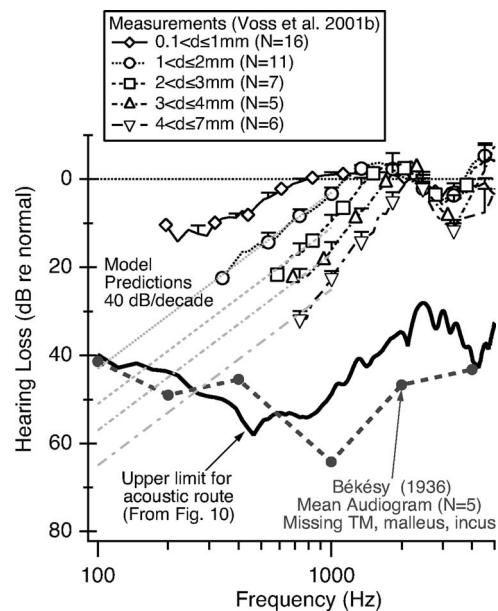
FIG. 11. "Hearing loss" (based on stapes-velocity V_S measurements in cadaveric temporal bones) resulting from perforations in the tympanic membrane. The upper five lines are mean measured loss in total transmission for five categories of perforation size (Voss *et al.*, 2001a, b). The curve labeled "Upper-limit for acoustic-route" is the transmission loss that would result if the acoustic route were dominant (from the "TM removed" curve in the lower plot of Fig. 10). Measurement of loss at the lower frequencies was not possible because the measurements were indistinguishable from noise. However, the model prediction for the loss increases at 40 dB/decade for low frequencies (Voss *et al.*, 2001c) as indicated by the straight gray lines for four perforation sizes. Also plotted is the mean audiogram collected by Békésy (1936; 1960 p. 105) on five subjects with missing TM, malleus, incus and stapes. The similarity of the two lower curves is consistent with the hypothesis that the acoustic route is the dominant path for sound transmission in these ears.

especially for the larger perforations, where the measured losses begin to approach the acoustic-route limit. The measured losses are limited at the lowest frequencies by the noise floor for stapes velocity, but models of middle-ear sound transmission with perforations predict that the low-frequency losses increase as frequency decreases (Voss *et al.*, 2001c); for perforations greater than about 1 mm in diameter, the model predicts that changes in the pressure difference across the TM dominate the total loss in the ossicular route and that at low frequencies these losses increase with decreasing frequency at 40 dB/decade. These predictions are included on the plot in Fig. 11, showing that the acoustic route could contribute substantially with the larger perforations at the lowest frequencies (e.g., less than 500 Hz).

Figure 11 also plots the mean audiogram measured by Békésy (1936) on five ears missing the TM, malleus, incus, and stapes. In these ears, it seems that the acoustic route could be responsible for hearing. That is, the measured hearing loss is comparable to the upper limit for the acoustic route for frequencies less than 4000 Hz. Thus, this audiogram is consistent with the hypothesis that hearing in this condition is determined by the acoustic route[5] (Peake *et al.*, 1992; Mehta *et al.*, 2006).

VI. CONCLUSIONS

Middle-ear sound transmission with perforated tympanic membranes is dominated by sound transmission via the os-

sicular system (ossicular route) for most perforation sizes; the acoustic route dominates when the TM is missing. Direct acoustic stimulation of the cochlear windows (acoustic route) provides a significant amount of sound transmission only at the lowest frequencies with a condition approaching total removal of the tympanic membrane. Sound transmission via the acoustic route is independent of perforation location but does depend on perforation size. These variations of sound transmission via the acoustic route result entirely from changes in transmission of sound from the ear canal to the middle-ear cavity, measured here as $H_{P_{CAV}}$. In fact, the dependence of the window-pressure ratio P_{RW}/P_{OW} and the window-pressure difference H_{WPD} on the pressure in the middle-ear cavity appear independent of (1) the presence of a tympanic-membrane perforation, (2) perforation size, and (3) perforation location. In other words, the relative magnitudes and angles of the pressures just outside the oval and round windows are unaffected by any tympanic-membrane perforation. The results (1) refute the common clinical view (described in the Introduction) that perforation location affects the relative phase of the sound pressures at the oval and round windows, and (2) provides the most complete existing description of the window pressure ratio P_{RW}/P_{OW} based on experimental data.

ACKNOWLEDGMENTS

[1]For the group of eight bones, there exist occasional frequency bands where the variation exceeds 0.25 dB in magnitude and 0.005 cycles in angle; such bands are rare and narrow. Fig. 3-3 of Voss (1998) provides an example.

[2]The plot of $\angle H_{P_{CAV}}$ for bone 24L with the TM intact shows an angle that is above zero at the lowest frequencies. However, the other ten bones have mean angles that are essentially zero at the lowest frequencies.

[3]Equation (9) is valid up to a frequency that depends on the volume of the middle-ear cavity. As the volume increases, the low-frequency range for which Eq. (9) holds decreases; plots for the Z_{CAV} corresponding to each preparation can be found in Fig. 2 of Voss et al. (2001c). Each f_{max} listed in Table II for $|H_{P_{CAV}}|$ in the range of 0.5–2.8 kHz generally occurs at a frequency where the approximation of Eq. (9) holds.

[4]Newton's second law in acoustics states: $\partial P/\partial x = -j\omega\rho V_x$. If the particle velocity $|V_x|$ going into the aditus is a maximum (i.e., at the resonant frequency), then $|\partial P/\partial x|$ will also be a maximum.

[5]These results of Békésy (1936) are unique in that they include thresholds for frequencies from 5 to 4000 Hz, which are shown in Fig. 11 only for 100–4000 Hz. The results below 100 Hz are difficult to interpret because the mean loss decreases for lower frequencies becoming about 10 dB at 5 Hz. These results led Békésy (1936) to conclude that "something apart from the pressure difference between the stapes and round window must be considered as producing the effects" (Békésy, 1960, p. 106). See Peake et al. (1992, pp. 258–259) for further discussion of this issue.

Békésy, G. v. (**1936**). "Zur Physik des Mittelohres und über das Hören bei fehlerhaftem Trommelfell (On the physics of the middle ear and hearing with a missing eardrum)," Akust. Z. **1**, 13–23 (English translation pp. 104–115 in Békésy, 1960).

Békésy, G. v. (**1947**). "The sound pressure difference between the round and the oval windows and the artificial window of labyrinthine fenestration," Acta Oto-Laryngol. **35**, 301–315.

Békésy, G. v. (**1960**). *Experiments in Hearing*, edited by E. G. Wever (McGraw–Hill, New York).

Efron, B., and Tibshirani, R. J. (**1993**). *An Introduction to the Bootstrap* (Chapman and Hall, London).

Gan, R. Z., Sun, Q., Feng, B., and Wood, M. W. (**2006**). "Acoustic-structural coupled finite element analysis for sound transmission in human ear—Pressure distributions," Med. Eng. Phys. **28**, 395–404.

Glasscock, M. E., and Shambaugh, G. E. (**1990**). *Surgery of the Ear*, 4th ed. (Saunders, Philadelphia).

Lynch, T. J. (**1981**). *Signal Processing by the Cat Middle Ear: Admittance and Transmission, Measurements and Models*, Ph.D. thesis, Massachusetts Institute of Technology.

Mehta, R. P., Rosowski, J. J., Voss, S. E., O'Neil, E., and Merchant, S. N. (**2006**). "Determinants of hearing loss in perforations of the tympanic membrane," Otol. Neurotol. **27**, 136–143.

Merchant, S. N., Ravicz, M. E., and Rosowski, J. J. (**1996**). "Acoustic input impedance of the stapes and cochlea in human temporal bones," Hear. Res. **97**, 30–45.

Merchant, S. N., Ravicz, M. E., Puria, S., Voss, S. E., Whittemore Jr., K. R., Peake, W. T., and Rosowski, J. J. (**1997a**). "Analysis of middle-ear mechanics and application to diseased and reconstructed ears," Am. J. Otol. **18**, 139–154.

Merchant, S. N., Ravicz, M. E., and Rosowski, J. J. (**1997b**). "Mechanics of type IV tympanoplasty: Experimental findings and surgical implications," Ann. Otol. Rhinol. Laryngol. **106**, 49–60.

Peake, W., Rosowski, J. J., and Lynch, T. J. (**1992**). "Middle-ear transmission: Acoustic versus ossicular coupling in cat and human," Hear. Res. **57**, 245–268.

Pickles, J. O. (**1987**). "Physiology of the ear," 5th ed., D. Wright, editor, in *Scott-Brown's Otolaryngology: Basic Sciences* (Butterworths, London), Vol. **1**, Chap. 2, pp. 59, 60.

Schmitt, H. (**1958**). "Uber die bedeutung der schalldrucktransformation und der schallprotektion für die hörschwelle (About the importance of sound pressure transformation and the acoustic protection for the hearing threshold)," Acta Oto-Laryngol. **49**, 71–80.

Schuknecht, H. (**1968**). "Temporal bone removal at autopsy," Arch. Oto-laryngol. **87**, 129–137.

Schuknecht, H. F. (**1993**). *Pathology of the Ear*, 2nd ed. (Lea and Febiger, Malvern, PA).

Stepp, C. E., and Voss, S. E. (**2005**). "Acoustics of the human middle-ear air space," J. Acoust. Soc. Am. **118**, 861–871.

Voss, S. E. (**1995**). *Is the Pressure Difference Between the Oval and Round Windows the Effective Acoustic Stimulus for the Inner Ear?* Master's thesis, Massachusetts Institute of Technology.

Voss, S. E. (**1998**). *Effects of Tympanic-Membrane Perforations on Middle-Ear Sound Transmission: Measurements, Mechanisms, and Models*, Ph.D. thesis, Massachusetts Institute of Technology.

Voss, S. E., Rosowski, J. J., and Peake, W. T. (**1996**). "Is the pressure difference between the oval and round windows the effective acoustic stimulus for the cochlea?" J. Acoust. Soc. Am. **100**, 1602–1616.

Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2000**). "Acoustic responses of the human middle ear," Hear. Res. **150**, 43–69.

Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2001a**). "How do tympanic-membrane perforations affect human middle-ear sound transmission?" Acta Oto-Laryngol. **121**, 169–173.

Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2001b**). "Middle-ear function with tympanic-membrane perforations. I Measurements and mechanisms," J. Acoust. Soc. Am. **110**, 1432–1444.

Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2001c**). "Middle-ear function with tympanic-membrane perforations. II. A simple model," J. Acoust. Soc. Am. **110**, 1445–1452.

Wever, E. G., and Lawrence, M. (**1954**). *Physiological Acoustics* (Princeton University Press, Princeton, NJ).

Sound pressure distribution and power flow within the gerbil ear canal from 100 Hz to 80 kHz

Michael E. Ravicz[a]
Eaton-Peabody Laboratory, Massachusetts Eye & Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

Elizabeth S. Olson
Department of Otolaryngology, Head and Neck Surgery, Columbia University, P & S 11-452, 630 West 168th Street, New York, New York 10032 and Biomedical Engineering Department, 1210 Amsterdam Avenue, New York, New York 10027

John J. Rosowski
Eaton-Peabody Laboratory, Massachusetts Eye & Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114, Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts 02115, and Harvard-MIT Division off Health Science and Technology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

Sound pressure was mapped in the bony ear canal of gerbils during closed-field sound stimulation at frequencies from 0.1 to 80 kHz. A 1.27-mm-diam probe-tube microphone or a 0.17-mm-diam fiber-optic miniature microphone was positioned along approximately longitudinal trajectories within the 2.3-mm-diam ear canal. Substantial spatial variations in sound pressure, sharp minima in magnitude, and half-cycle phase changes occurred at frequencies >30 kHz. The sound frequencies of these transitions increased with decreasing distance from the tympanic membrane (TM). Sound pressure measured orthogonally across the surface of the TM showed only small variations at frequencies below 60 kHz. Hence, the ear canal sound field can be described fairly well as a one-dimensional standing wave pattern. Ear-canal power reflectance estimated from longitudinal spatial variations was roughly constant at 0.2–0.5 at frequencies between 30 and 45 kHz. In contrast, reflectance increased at higher frequencies to at least 0.8 above 60 kHz. Sound pressure was also mapped in a microphone-terminated uniform tube—an "artificial ear." Comparison with ear canal sound fields suggests that an artificial ear or "artificial cavity calibration" technique may underestimate the *in situ* sound pressure by 5–15 dB between 40 and 60 kHz.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2769625]

I. INTRODUCTION

A. Spatial variations in sound within the ear canal

This paper is part of a study to investigate high-frequency sound transmission through the external and middle ear of the Mongolian gerbil *Meriones unguiculatus* and has ramifications for similar investigations in humans and other species. In studies of auditory function, the auditory input is generally taken as the sound pressure in the ear canal (EC) just lateral to the tympanic membrane (TM), and explorations of auditory function at any point along the auditory pathway (middle ear, cochlea, auditory nerve and beyond) generally require an accurate assessment of this input. At low frequencies, sound pressure is approximately uniform throughout the ear canal, and auditory input can be estimated accurately from a sound pressure measurement anywhere within the ear canal. At high frequencies, however, sound pressure can vary considerably within the ear canal and across the TM, and accurate measurement of auditory input is problematic.

At such high frequencies, spatial variations in the sound field related to the wave nature of sound can be described in terms of modal patterns with local maxima and minima in sound pressure magnitude and differences in phase that occur along the length or across the diameter of the ear canal. Sound pressure variations along the ear canal length L can become significant at frequencies where the sound wavelength $\lambda < \sim 10L$.[1] Variations across the ear canal diameter d can become significant near changes in ear canal cross-section or mechanical properties (such as near the TM; e.g., Rabbitt and Holmes, 1988). The first mode that creates these transverse variations can propagate along the ear canal at frequencies where $\lambda < d/0.59$ (Morse and Ingard, 1968, p. 511; Stinson and Khanna, 1994), lower if the ear canal cross section is noncircular. The mechanisms of these frequency-dependent spatial variations in ear-canal sound pressure are discussed in more detail in Sec. II.

Such spatial variations in EC sound pressure are usually significant at audible frequencies. For example, for the hu-

---
a)Electronic mail: mike_ravicz@meei.harvard.edu

man ear canal with length $L=22.5$ mm (Shaw, 1974) and EC area $\approx$TM area$=60$ mm^2 (Wever and Lawrence, 1954), λ $<10L$ above 500 Hz, $d \approx 8.7$ mm, and $\lambda < d/0.59$ above 24 kHz. Considering that the range of human hearing is 20 Hz–20 kHz (Sivian and White, 1933), it is clear that longitudinal sound pressure variations exist in the ear canal well below the upper frequency limit of hearing. The frequency range of gerbil hearing has been described as 210 Hz–46 kHz (threshold within 30 dB of best threshold; Rosowski, 1992). In a gerbil preparation in which the cartilaginous ear canal has been removed (as in many physiological experiments), the length of the remaining bony ear canal is about 4 mm. The EC diameter is 2.3 mm near its midpoint but about 5 mm near the TM [mean of Lay (1972) and Ravicz (1990)]. With these dimensions, $\lambda < 10L$ above 8 kHz and $\lambda < d/0.59$ near the TM above 40 kHz. Therefore, substantial sound pressure variations could exist both along the EC length and across the EC diameter at frequencies below the 46-kHz upper limit of gerbil hearing. Similarly, sound pressure can vary throughout the ear canal of other species within their audible range.

Several techniques have been developed to estimate the middle-ear input at high frequencies. One common approach in nonhuman ears is to use a probe-tube microphone to measure the sound pressure at a point just lateral to the TM. This method is difficult both in a clinical setting as well as in small animals with narrow ECs. In some animal experiments, a portion of the ear canal is removed to shorten the effective length of the ear canal between the measurement location and the TM (as described earlier for gerbil). A second approach is to use measurements of the sound pressure in a convenient location relatively distant from the TM to estimate the sound pressure at the TM using a mathematical model of the ear canal (e.g., Stevens et al., 1987; Huang et al., 2000). A third and probably the most common approach for studies of human hearing is to calibrate the earphone output in an "artificial ear" that mimics the essential dimensions of the ear canal as well as the load impedance of the TM on the ear canal. Such an approach has been used to calibrate audiometric earphones (Zwislocki et al., 1988).

A variation of this third approach has been used in recent studies of high-frequency hearing and auditory mechanics in small mammals (Pearce et al., 2001; Overstreet and Ruggero, 2002; Overstreet et al., 2003). In these studies, the stimulus earphone was calibrated against a reference microphone that was coupled to the earphone via an excised ear canal or a short tube of dimensions similar to the ear canal—an "artificial cavity calibration." There are problems with this approach: In particular, the acoustic impedance of the reference microphone terminating the artificial ear is a high-magnitude stiffness that is unlikely to be representative of the impedance of the intact TM. While the impedance at the TM is not well described in any animal at frequencies above about 20 kHz, various authors have suggested that the high-frequency TM impedance in cats is resistive and approximately matched to the characteristic impedance of the ear canal [Lynch, 1981; Lynch et al., 1994; inferred from Allen (1986) and Puria and Allen (1998)]. It should be noted, however, that in none of these studies did the frequency

range of measurement approach the upper frequency limit of hearing. If the terminating impedance is larger than that of the TM, the sound field within the ear canal will be altered from the natural case, and the calibration may be biased, as we demonstrate in this paper.

B. Sound field measurements in gerbil ear canals

The topics of variations in EC sound pressure and EC sound pressure calibration difficulties at medium to high frequencies have been addressed previously for humans and cats for frequencies as high as 20 kHz (Stinson et al., 1982; Lawton and Stinson, 1986; Khanna and Stinson, 1985; Stinson and Khanna, 1994). Here we extend the discussion to higher frequencies in the Mongolian gerbil, a common subject for auditory measurements whose ear canals are smaller than those of cat or human. The frequency range of measurements includes the entire range of gerbil hearing.

This paper contains measurements of spatial variations in sound pressure within the gerbil ear canal near the TM. Sound pressure variations were also measured in a cylindrical artificial ear terminated by a reference microphone. Measurements were obtained from two different laboratories using two different sound-pressure measuring devices of different size and sensitivity: a probe-tube microphone, and a fiber-optic miniature pressure sensor (Olson, 1998), which acts as a miniature microphone in air. Most of the probe-tube measurements in ear canals and in the artificial ear were performed in the Eaton-Peabody Laboratory at the Massachusetts Eye and Ear Infirmary (MEEI), while the fiber-optic miniature microphone measurements in ear canals were performed at Princeton University (early experiments) and the Fowler Laboratory at Columbia University (later experiments). The more extensive ear-canal sound pressure maps were made with the probe-tube system; results with the smaller fiber-optic microphone confirm those findings and establish that the probe tube did not affect the sound field significantly.

Our measurements address the following questions: (1) How does the sound pressure vary in the terminal portion of the gerbil ear canal? (2) Do sound-field measurements in a calibration cavity provide an accurate estimate of the sound pressure near the umbo at the center of the pars tensa of the TM? (3) What fraction of the sound energy in the ear canal is reflected at the TM at different frequencies? Our findings can be generalized to other species by a simple scaling, using the anatomical dimensions for the gerbil ear canal provided here.

II. THEORY

A. Sound pressure distribution and standing waves in a model ear canal

For the purposes of the analysis in this paper, it will be convenient to consider the ear canal as a uniform cylindrical tube of length L and diameter d. The tube is coupled to a sound source at one end ($x=-L$) and terminated with the TM at the other end ($x=0$). Consider a source-generated forward-traveling uniform plane sound wave $p_i(x,t)$ (fundamental mode) with frequency f described by $p_i(x,t)=A \cos(\omega t - kx)$, where $\omega = 2\pi f$, t is time, k is the wave number$=2\pi/\lambda$, and x

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Ravicz et al.: Gerbil ear canal sound pressure distribution   2155

is the longitudinal dimension. At the termination (the TM), part of the incident (forward-traveling) wave is absorbed and part is reflected. The reflected wave $p_r(x,t)$ travels backward along the ear canal toward the source (in the $-x$ direction) and is described by $p_r(x,t) = B\cos(\omega t + kx + \theta)$, where θ is the phase difference at the TM between the incident and reflected waves. The resulting sound field in the ear canal is the sum of the sound pressures of the incident and reflected waves:

$$p(x,t) = p_i(x,t) + p_r(x,t)$$
$$= A\cos(\omega t - kx) + B\cos(\omega t + kx + \theta). \quad (1)$$

At frequencies where λ is greater than $10L$, spatial variations in sound pressure within the ear canal are small, and the ear canal can be considered a lumped element. However, at higher frequencies (shorter wavelengths), sound pressure will vary substantially longitudinally along the ear canal (e.g., Kinsler *et al.*, 1982). Similarly, at frequencies where $\lambda < d/0.59$, the sound pressure can vary substantially transversely across the EC diameter; these transverse variations (higher-order modes) will be considered later.

The relative magnitude B and relative phase of the reflected wave θ are determined by the nature of the EC termination. If the TM is perfectly reflecting, such that all sound power incident on the TM is reflected backward toward the source, then $A=B$, and the sound field in the ear canal is a standing wave in which the amplitude varies spatially as $|\mathbf{P}| = 2A\cos(kx + \theta/2)$. (If the TM were not only perfectly reflecting but also perfectly rigid, $\theta=0$.) If all of the sound power incident on the TM is absorbed (transmitted into the middle ear), there is no reflected wave, $B=0$, and the sound field is described by the forward traveling wave $A\cos(\omega t - kx)$ in which the sound amplitude$=A$ for all x throughout the ear canal and the phase increases as kx. If the incident wave is partially absorbed and partially reflected, the EC sound field has features of both traveling and standing waves (e.g., Kinsler *et al.*, 1982).

Figure 1(a) illustrates how the longitudinal distribution of ear canal sound pressure magnitude $|\mathbf{P}|$ and phase $\angle\mathbf{P}$ varies in a model ear canal for different values of B/A, the amount of reflection at the TM (pressure reflection coefficient). In this model, the EC length is 5 mm and the EC diameter is small enough that we need consider only uniform plane waves. Values of B/A between 0 and 0.9 correspond to varying the magnitude of the TM impedance between the characteristic impedance of the tube $Z_0 = \rho_0 c/S$ and about $20 \times Z_0$ (nearly rigid). [c is the propagation velocity of sound in air, ρ_0 is the density of air, and S is the EC cross-section area$= \pi d^2/4$, e.g., Kinsler *et al.*, (1982)]. Curves were generated from

$$|\mathbf{P}(x)| = [1 + (B/A)^2 + 2(B/A)\cos(2kx - \theta)]^{1/2} \quad (2a)$$

$$\angle\mathbf{P}(x) = \frac{\theta}{2} + \arctan\left[\frac{1 - (B/A)}{1 + (B/A)}\tan\left(kx - \frac{\theta}{2}\right)\right] \quad (2b)$$

[from Stinson, (1985a), Eqs. (30) and (32)]. The magnitude of each curve in Fig. 1(a) is normalized by the maximum
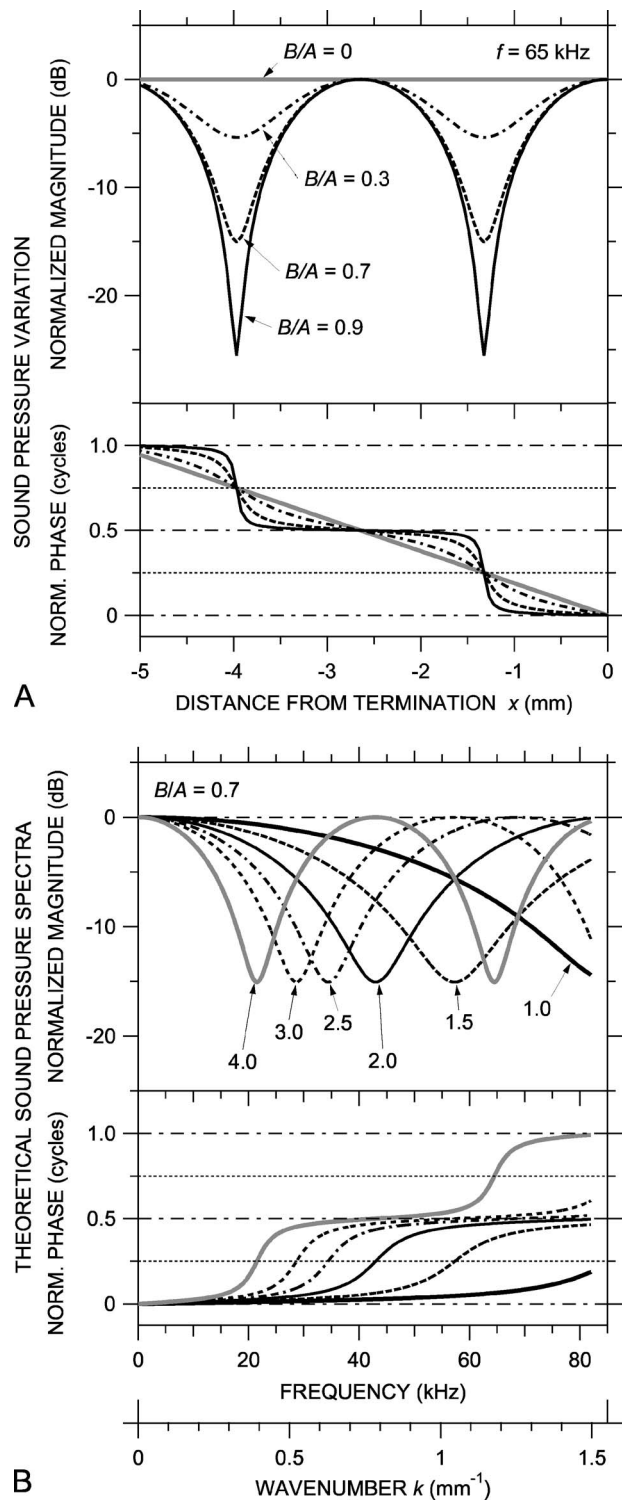
FIG. 1. Theoretical standing waves in a uniform cylindrical tube. (A) Spatial variations in sound pressure magnitude (top) and phase (bottom) at a given frequency for different values of B/A, the ratio of the amplitude of a sound wave reflected from the termination to the incident sound wave amplitude. (B) Computed broadband sound pressure spectra at different locations for $B/A=0.7$. If B/A changes slowly with frequency, the spectra are equivalent to the spatial sound pressure distribution—see lower horizontal scale in wave number $k=2\pi f/c$.

sound pressure achievable within the tube $(B+A)$, and the phase is normalized by the phase at the termination.[2] The sound frequency is 65 kHz, so the wavelength is 5.3 mm. For simplicity, we set $\theta=0$.

Ravicz *et al.*: Gerbil ear canal sound pressure distribution

When $B/A=0$ [zero reflection, total transmission; the thick gray line in Fig. 1(a)], the sound pressure amplitude is uniform throughout the tube and the phase of the sound varies linearly with distance from the TM ($\angle \mathbf{P}$ slope is constant with position). The sound field in the ear canal is a uniform traveling wave.

For nonzero reflection (e.g., $B/A=0.3$, dot-dashed line), broad local maxima and narrow local minima (notches) appear in the spatial $|\mathbf{P}|$ distribution, and $\angle \mathbf{P}$ shows slight increases and decreases from the constant slope of the zero-reflection case. The local $|\mathbf{P}|$ maxima occur where the incident and reflected waves combine constructively; this is also where the $\angle \mathbf{P}$ slope is shallowest and equal to 0, 0.5, or 1 cycle, in this model ear canal at $x=0$ and -2.65 mm. The $|\mathbf{P}|$ notches occur where the incident and reflected waves combine destructively, in this model ear canal at $x=-1.33$ and -3.98 mm, where $x=\lambda/4$ (a "quarter-wave resonance") and $3\lambda/4$. These locations are also where the $\angle \mathbf{P}$ slope is steepest, where $\angle \mathbf{P}=0.25$ or 0.75 cycles. To state the rule, $|\mathbf{P}|$ notches occur and $\angle \mathbf{P}=(2n-1)/4$ cycles where $kx=\pi/2, 3\pi/2, 5\pi/2,\ldots,(2n-1)\pi/2$ [after Stinson (1985a)].

As B/A increases toward 1 [in our model ear canal, as the magnitude of the TM impedance increases further; Fig. 1(a), dashed and solid curves], the $|\mathbf{P}|$ notches deepen and the spatial $\angle \mathbf{P}$ distribution becomes more step-like. If $B/A=1$ (total reflection, zero transmission), $|\mathbf{P}|$ is zero at the minima, and the sound field in the ear canal is a pure standing wave. For $0<B/A<1$, the sound field can be thought of as a combination of a traveling wave and a standing wave: Some of the incident wave is transmitted into the TM and some is reflected to form a pattern of $|\mathbf{P}|$ maxima and minima and associated $\angle \mathbf{P}$ changes.

In this uniform ear canal, sound power reflectance at the TM $\mathfrak{R}$ is related to the standing wave ratio (SWR), the ratio of $|\mathbf{P}|$ maximum to minimum, by

$$\mathfrak{R}=(B/A)^2=[(\mathrm{SWR}-1)/(\mathrm{SWR}+1)]^2 \qquad (3)$$

(e.g., Magnusson, 1965; Kinsler et al., 1982; Stinson et al., 1982).

In addition to the fundamental-mode uniform plane waves discussed earlier, higher-order-mode nonuniform plane waves may also be present in a tube and may propagate down the ear canal. These nonuniform plane waves reflect off the tube or EC walls and cause transverse variations in sound-pressure magnitude and phase across the tube or EC diameter (Morse and Ingard, 1968, p. 511; Fletcher, 1992, p. 198). Higher-order modes may be induced by a nonuniformity in the source generating the wave, the reflecting termination, or the tube itself. A large potential source for transverse variations in ear canal sound pressure is the tympanic membrane, which is known to vibrate in complicated patterns (e.g., Khanna and Tonndorf, 1972; Rabbitt and Holmes, 1988; Furlong et al., 2006; Rosowski et al., 2006). At frequencies where $\lambda>d/0.59$, any nonuniform plane wave that is generated by the TM will decay exponentially[3] within a short distance of the TM. For example, in the terminal portion of the gerbil ear canal at frequencies less than 37 kHz, the lowest nonuniform mode generated at the TM will decay by more than 3 dB at 0.5 mm from the TM and by 20 dB at

3 mm from the TM. At higher frequencies, nonuniform modes do not decay but propagate back up the ear canal. The resultant EC sound field in the ear canal is the sum of incident and reflected waves of all modes present.

Such transverse spatial sound pressure variations complicate the estimation of power reflection in the ear canal: Though the notches in the uniform-plane-wave standing wave pattern in a tube or ear canal with a reflective termination may be quite deep, the sound pressure associated with higher-order modes may fill in the deepest part of the notches, thereby decreasing their apparent depth and causing the SWR and hence the power reflectance to be underestimated. Higher-order modes may also cause phase irregularities in the standing wave pattern. (Microphone or environmental noise could have a similar effect.) In this way, higher-order modes limit the maximum SWR and ear canal reflectance that we could measure.

B. Sound pressure spectra at various locations in the model ear canal

Because the spatial term of the wave equation [Eq. (1)] depends on both location x and sound frequency f (via the wave number $k=2\pi f/c$), we can prepare a "dual" of the standing wave pattern in Fig. 1(a) (which shows the sound pressure at many locations in the ear canal at one frequency) by plotting the sound pressure at one location at many frequencies (spectrum). Figure 1(b) shows such dual sound pressure spectra (magnitude and phase) for $B/A=0.7$ at six locations in the model ear canal (at 1, 1.5, 2, 2.5, 3, and 4 mm from the TM). The common feature in Figs. 1(a) and 1(b) is the pattern of repeated magnitude notches and 0.5 cycle phase changes. Such simple relationships between the spatial and frequency dependence will occur as long as B/A varies slowly with frequency. The lower abscissa scale in Fig. 1(b), which scales the patterns in k, emphasizes this duality.

III. METHODS

A. Animal preparation

1. At Massachusetts Eye and Ear Infirmary (MEEI)

The preparation of the experimental animals was approved by the MEEI Institutional Animal Care and Use Committee and is similar to that described previously (e.g., Ravicz et al., 1992). Thirteen young adult Mongolian gerbils were used. The animals were anesthetized with an intraperitoneal injection of sodium pentobarbital (Nembutal, 40 mg/kg) together with intramuscular Ketamine injections (25 mg/kg). A tracheal cannula was inserted. The soft tissue was removed from the surface of the skull around the opening of the bony ear canal. Two or three small holes were made in the wall of the auditory bulla to preclude a buildup of static pressure within the middle-ear cavity and to provide visual access to the stapes.

The intact gerbil ear canal is long, curved, and narrow and flares near the TM, which has a diameter approximately twice that of the ear canal (5 mm vs 2.3 mm; see previous text) and is oriented nearly parallel to the axis of the more distal ear canal (Fig. 2). Consequently, bringing a probe tube

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Ravicz et al.: Gerbil ear canal sound pressure distribution   2157
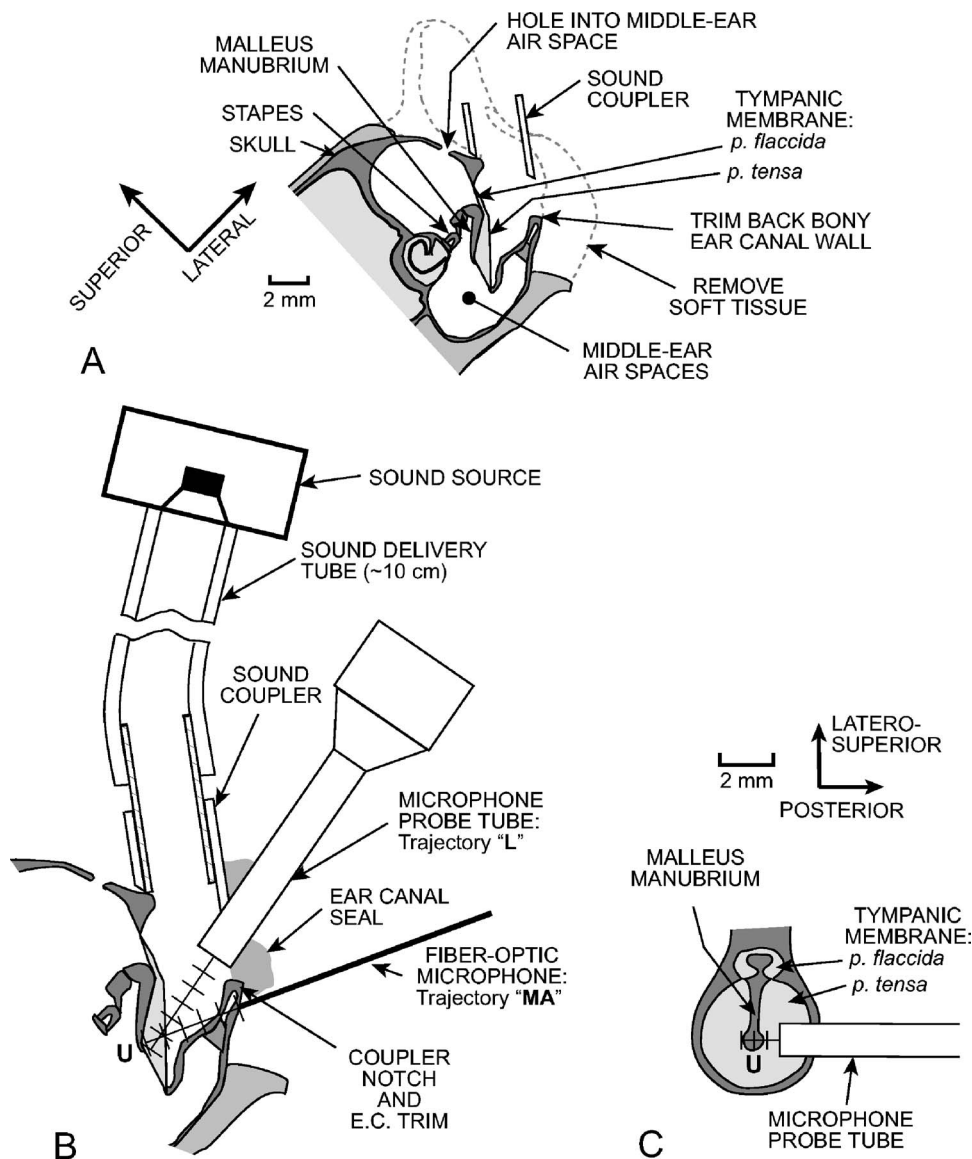
FIG. 2. Schematic section through a gerbil head showing preparation methods and measurement trajectories. (A) Coronal section. The external ear and soft tissue were removed, the lateral wall of the bony ear canal (EC) opening was trimmed back, and a brass tube sound coupler was placed at the opening of the bony ear canal. A reflector was placed on the stapes posterior crus, and the laser vibrometer beam was shined through one or more holes in the middle ear wall toward the reflector. (B) Details of the coronal EC section showing EC sound pressure (P_{EC}) measurement trajectories. The microphone probe tube (MEEI) or fiber-optic microphone (Princeton and Columbia) was introduced through a notch in the coupler wall and advanced in 0.5-mm steps toward the umbo "U" along an approximately longitudinal trajectory "L." P_{EC} was also measured with the fiber-optic microphone along an medio-axial trajectory through the middle-ear wall "MA" toward the umbo. (C) Section perpendicular to the coronal section in panel (B). P_{EC} was also measured along a transverse trajectory across the ear canal over the umbo approximately 0.5 mm from the tympanic ring.

near the TM is difficult. To shorten the ear canal, the pinna flange and cartilaginous EC were removed, and as much as possible of the ventro-lateral wall of the bony EC opening was removed without breaking into the middle-ear cavity. A short cylindrical brass tube [3.2 mm inside diameter (i.d.) $\times 3-5$ mm length] was cemented around the bony EC opening with dental cement or cyanoacrylate glue [Fig. 2(a)]. The tube permitted reliable coupling and recoupling of the ear to the high-frequency sound source [Fig. 2(b)]. In earlier experiments the sound coupler was left square-ended and covered part of the opening in the lateral EC wall. In the six later experiments, in which most of the data presented here were taken, the coupler end was beveled and notched to widen the opening between the coupler and the lateral-wall remnant [Fig. 2(a)].

2. At Princeton and Columbia

The preparation of the experimental animals was approved by the Institutional Animal Care and Use Committees of Princeton University and Columbia University as applicable. Anesthesia and animal monitoring were as described

in Olson (1998) except that 40 mg/kg Ketamine was administered preoperatively instead of acepromazone.[4] Three young adult Mongolian gerbils were used; one was just postmortem when measurements commenced.

B. Experimental methods

The experiments were conducted in a sound-isolation booth. Heart rate was monitored throughout the experiment, and body temperature was maintained at 37 ± 1 °C with a heating pad. Anesthesia boosters at half the initial surgical dose were provided as necessary to suppress the withdrawal reflex. The animal's head was held securely in a custom snout clamp. At the conclusion of the live experiment the animal was euthanized with a fatal dose of anesthesia.

Sound was produced by an electrostatic earphone (TDT EC1 and amplifier ED1, Tucker-Davis Technologies, Alachua, FL) and delivered to the ear canal via a flexible sound delivery tube (Tygon) approximately the same i.d. (2.4 mm) as the ear canal and about 10 cm in length [Fig. 2(b)], with a short brass section that allowed the sound delivery tube to be removed and replaced repeatedly. Sound pressure P_{EC} was

Ravicz *et al.*: Gerbil ear canal sound pressure distribution

measured at various locations within the remaining ear canal with the measurement microphone [the probe-tube microphone (MEEI) or the miniature fiber-optic microphone (Princeton/Columbia)] mounted in a micromanipulator (Narishige or Marzhauser). Both the magnitude and phase of $\mathbf{P}_{EC}$ relative to the stimulus were measured. The probe tube was introduced into the ear canal through the notch in the sound coupler, and $\mathbf{P}_{EC}$ was measured in various locations as the probe tube was advanced and retracted in 0.25–0.5 mm steps along an approximately longitudinal trajectory down the ear canal toward the umbo [trajectory "L" in Fig. 2(b)]. Measurements were made with the gap between the probe tube and the sound coupler left unsealed and with the gap sealed with saturated cotton, which was effectively acoustically rigid. The probe tube was advanced until changes in microphone output and stapes velocity (see the following) suggested that it had touched the TM. (Sound pressure measured at this position is not included.) The innermost probe tube position was confirmed by viewing scuff marks on the TM through an operating microscope after the probe tube had been withdrawn.

In two early experiments (at Princeton), the miniature fiber-optic microphone was introduced into the ear canal via natural cracks in the bone of the EC lateral wall (Olson, 1998) and advanced toward the umbo in an approximately medio-axial direction intermediate between longitudinal and horizontal [trajectory "MA" in Fig. 2(b)]. In the third experiment (at Columbia), the MEEI sound coupler was used, and the microphone was introduced and advanced approximately longitudinally along trajectory L as described earlier. In all experiments, the microphone was advanced directly toward the umbo, which was visible throughout the experiment. Low-frequency fluctuations in the fiber-optic microphone output signaled that the microphone tip had touched the TM. The third experiment was performed just post-mortem.

Stapes velocity was measured simultaneously with $\mathbf{P}_{EC}$ in the longitudinal-trajectory MEEI experiments to evaluate the effect of the probe tube on the sound field within the ear canal. The velocity of small reflectors placed on the posterior crus of the stapes was measured with a laser-Doppler vibrometer (OPV 501/2600, Polytec PI, Waldbronn, Germany) focused through one of the bulla holes (usually posterior to the ear canal—Rosowski et al., 1999; Ravicz and Rosowski, 2004). Stapes velocity was also used to monitor the position of the probe tube: When the probe-tube tip touched the TM, stapes velocity decreased at low frequencies.

In three ears, $\mathbf{P}_{EC}$ was also measured at various points spaced at 0.5 mm along a transverse trajectory [Fig. 2(c)] from the ear canal wall just lateral to the posterior tympanic ring to a point in the ear canal just lateral to the anterior tympanic ring. This trajectory passed approximately 0.5 mm lateral to the umbo. For these measurements, a hole was chipped in the posterior wall of the ear canal just lateral and posterior to the tympanic ring [Fig. 2(c)]. This access was obstructed by the animal's shoulder in intact animals, so these measurements were performed post-mortem after separating the head from the body.

C. Stimulus generation and response measurement

1. At MEEI

In the MEEI experiments, custom software (written in LABVIEW 7.1, National Instruments, Austin, TX) on a desktop computer synthesized stimuli, controlled equipment, acquired and averaged responses synchronously with the stimulus, and computed the transfer function between the response (microphone or laser-Doppler vibrometer) and the stimulus. Acoustical and mechanical responses were produced by sound stimuli delivered to the earphone via a signal generator (HP 33120A, Hewlett-Packard, Palo Alto, CA). The stimulus was a computer-generated broadband chirp with evenly spaced components from 50 Hz to 99 kHz. For some measurements, the spectrum of the chirp stimulus was modified to reduce the level at frequencies below 5 kHz by 20 dB, so as to produce more equal stimulus sound pressures over a broader frequency range. Stimulus levels were set by a computer-controlled attenuator (TDT PA5) and the output amplifier stage of a reconstruction filter (Krohn-Hite 3901, Avon MA). Responses were digitized synchronously to the stimulus signal by a National Instruments PCI-4451 A-D card triggered by the signal generator.

2. At Princeton and Columbia

The stimulus generation and response measurement system was similar but made use of a sequence of pure tones (step size 10–15 pts./octave at Princeton, 500 Hz at Columbia). The responses to the tones were digitized synchronously with signal generation by a paired D-to-A/A-to-D converter (Princeton: TDT System II DD1; Columbia: TDT System 3 RX6; includes attenuator and driver amplifier). Sound stimuli in early experiments were delivered to the ear canal by a dynamic earphone (based on Radio Shack 40-1377, Ft. Worth, TX).

D. Probe tube and miniature fiber-optic microphone design and calibration

1. Probe-tube microphone (at MEEI)

The probe-tube microphone used at MEEI comprised a $\frac{1}{2}$ in. condenser microphone (2530, Larson-Davis, Provo, UT) and a short (13 mm) probe tube constructed of commercially available hypodermic tubing [18-gauge thin-wall, 0.97 mm i.d., 1.27 mm outer diameter (o.d.)] coupled to the microphone via a short (7 mm) exponential horn and phase plug. Experimentation showed that the 0.97 mm i.d. was sufficient to transmit sound up to 80 kHz without unacceptably large viscous losses. The receptive field of the probe tube microphone extends approximately 0.3–0.4 mm beyond its tip ($\sim 0.6 \times$ the probe tube radius; Ando, 1968; Beranek, 1986).

2. Miniature fiber-optic microphone (at Princeton and Columbia)

The miniature fiber-optic microphone has been described previously (Olson, 1998). It consists of a glass capillary of 170 μm o.d., tipped with a gold-coated polymer diaphragm that is a fraction of a micrometer thick. Light
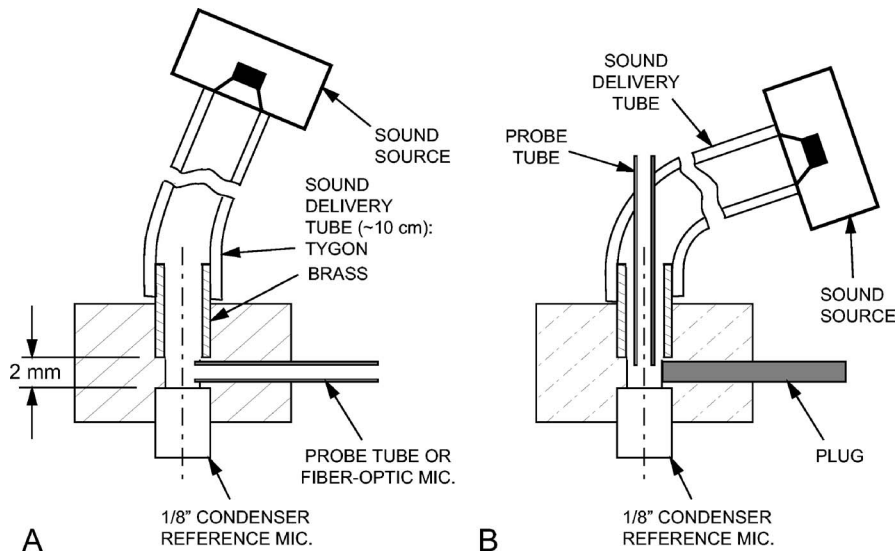
FIG. 3. Section through the calibration coupler used (A) for microphone calibrations and (B) as an "artificial ear." The sound delivery tube was 2.4-mm i.d. Tygon with a 2.4 mm i.d.×6 mm long brass section that fit into the acrylic body of the calibration coupler (2.3 mm diameter in front of the $\frac{1}{8}$-in. reference microphone). For the "artificial ear calibration," the setup in (A) was used, the probe tube was replaced with a plug as in (B), and the brass tube was withdrawn about 3 mm from the position shown.

from an infrared light-emitting diode is delivered via an optical fiber inserted into the capillary and reflects from the diaphragm. The amount of light returning through the optical fiber to a photodetector varies linearly with the pressure-induced motion of the diaphragm (Hu *et al.*, 1992). The receptive area of the diaphragm is 100 μm in diameter, so the receptive field extends less than 0.1 mm from the diaphragm.

3. Calibration coupler and reference microphone

Both microphones were calibrated in a custom calibration coupler that positioned the measurement microphone at grazing incidence within 1 mm of the diaphragm of an $\frac{1}{8}$ in. reference pressure microphone (G.R.A.S. 40-DP, N. Olmstead OH) of known sensitivity—see Fig. 3(a). The i.d. of the coupler near the reference microphone was 2.3 mm. The tip of the measurement microphone protruded approximately 0.25 mm past the wall. The calibration coupler was also used as an artificial ear [Fig. 3(b)] and for an artificial cavity calibration—discussed in Sec. IV D.

4. Microphone calibrations

The microphone calibrations are shown in Fig. 4. The magnitude and phase of the probe-tube microphone frequency response (solid line) varied substantially with frequency. The frequencies of the magnitude peaks and the phase accumulation (equivalent to a 60-μs delay) are generally consistent with the 20-mm length of the probe-tube assembly. When the probe-tube tip was plugged, the probe-tube microphone output decreased by at least 30 dB (except at isolated frequencies near 60 kHz, where the reduction was 15–20 dB), which implies that probe-tube acoustic artifact (sound conducted through the probe tube walls) was negligible. The microphone noise floor was 5–10 dB below the acoustic artifact.

Because the fiber-optic microphone has no probe tube and employs a very small receptive membrane, it is expected to have a flat frequency response to high frequencies. Its response in Fig. 4 (dashed line) was quite flat to 50 kHz but displayed a gradual decrease in sensitivity at the higher frequencies that reduced the response ~6 dB at ~70 kHz. The

fiber-optic microphone had a higher noise floor than the probe tube microphone; its output for P_{EC} measurements was generally only ~20 dB above its noise floor.

Longitudinal and transverse variations in the sound field within the calibration coupler and across the diaphragm of the reference microphone impose a high-frequency limit on the accuracy of the microphone calibrations. Differences in sound pressure between the measurement microphone location and the reference microphone due to longitudinal waves were expected to be less than 2 dB below 35 kHz, <6 dB below 60 kHz, and ~12 dB at 80 kHz with the measurement microphone located 1 mm from the nearly rigid reference
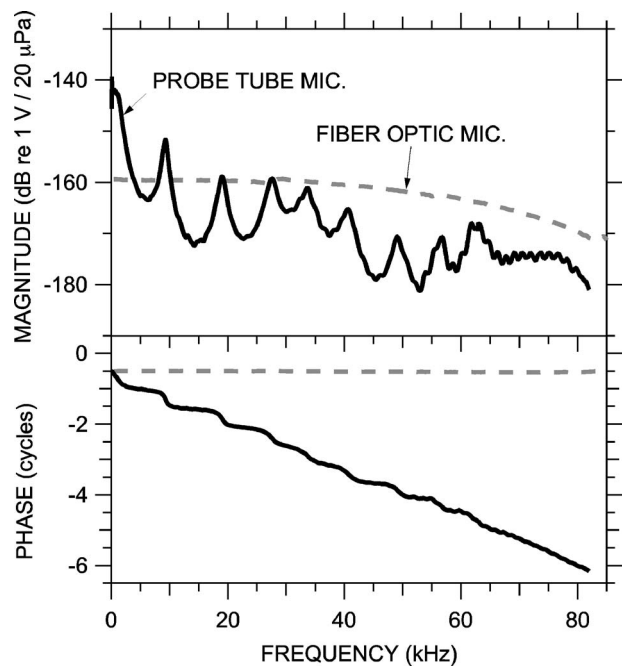
FIG. 4. Sensitivity and frequency response of the probe-tube microphone (MEEI, solid curve) and miniature fiber optic microphone (Princeton and Columbia, dashed curve). Top: Magnitude; bottom: phase relative to sound pressure condensation.

microphone diaphragm.[5] These differences limit the frequency range of absolute sound pressure measurements (Fig. 14) but do not affect ratios of ear canal sound pressure (Figs. 5–13).

Transverse variations in sound pressure across the coupler diameter due to higher-order modes needed to be taken into account because the measurement location is near their source at the reference microphone diaphragm (e.g., Keefe and Benade, 1981). The position of the sound delivery tube along the coupler axis and the close match between the coupler and sound delivery tube i.d.s minimized antisymmetric and axisymmetric higher-order modes, respectively. Differences in sound pressure across the coupler diameter were generally antisymmetric: <3 dB for the fiber-optic microphone, <5 dB below 68 kHz for the probe tube microphone. For calibration, we computed the area-averaged sound pressure (sum of pressure measured at different radial points normalized by the area of an annulus at that radius), as suggested in Keefe and Benade (1981). (The reference microphone output follows the average pressure over its area.)

IV. RESULTS

The MEEI experiments are emphasized because they included more experimental animals and approach angles, they sampled the frequency very finely (50 Hz), and the signal-to-noise ratio was higher; unless noted, results are from MEEI. The fine frequency sampling is particularly important to computations of sound reflection. The Princeton and Columbia results using the miniature fiber-optic microphone are shown in three animals and primarily provide confirmation of the MEEI results. A possibly important difference between the MEEI and Princeton/Columbia experimental methods was the size of the pressure measuring device: the 1.27-mm-diam probe-tube microphone was larger relative to the 2.3-mm-diam ear canal than the 170 μm-diam fiber-optic microphone. The similarity of the results with the two methods is consistent with the larger probe tube having no significant effect on the sound field.

A. Sound field longitudinally within the bony ear canal

The sound pressure $\mathbf{P}_{EC}$ within the bony terminal portion of the ear canal was measured approximately longitudinally along the ear canal by advancing the probe tube along trajectory L from the lateral wall of the ear canal to a point close to the umbo, as shown in Fig. 2(b). The receptive field of the probe tube microphone extends 0.3–0.4 mm beyond the tip (see Sec. III), so for measurements of the sound field near the umbo $\mathbf{P}_U$ (0-mm reference location), the probe tip was actually approximately 0.5 mm from the umbo. Because we are interested in spatial variations in $\mathbf{P}_{EC}$, we examine the ratio of $\mathbf{P}_{EC}$ measured at various locations along the trajectory to $\mathbf{P}_U$. Such ratios are denoted as $\Delta\mathbf{P}_{EC}$. The top panel of Fig. 5(a) shows the magnitude of the spatial variations in sound pressure $|\Delta\mathbf{P}_{EC}|$ in an ear canal (0402R); the bottom panel shows the corresponding variations in phase $\angle\Delta\mathbf{P}_{EC}$. Microphone and physiological noise were at least 30 dB be-

low $|\Delta\mathbf{P}_{EC}|$ maxima and about 10 dB below $|\Delta\mathbf{P}_{EC}|$ minima. At frequencies below 30 kHz, the variations in $\Delta\mathbf{P}_{EC}$ along the trajectory were less than 6 dB in magnitude and 0.2 cycles in phase. The sound pressure was lowest near the umbo: $|\Delta\mathbf{P}_{EC}|$ was generally greater than 0 dB. Below 10 kHz the phase differences were less than 0.1 cycle; between 10 and 30 kHz, $\angle\Delta\mathbf{P}_{EC}$ became more negative as the distance from the umbo increased, consistent with sound wave propagation along the trajectory. Above 30 kHz, $|\Delta\mathbf{P}_{EC}|$ varied between +10 and −25 dB and $\angle\Delta\mathbf{P}_{EC}$ varied up to 1 cycle. At this 30-kHz break frequency, the 3-mm distance between the EC wall and the reference location near the umbo corresponds to about one-quarter of the sound stimulus wavelength.

These high-frequency variations in $\Delta\mathbf{P}_{EC}$ magnitude and phase can be seen more clearly when each $\Delta\mathbf{P}_{EC}$ measurement is plotted separately on a linear frequency scale. Figure 5(b) shows $\Delta\mathbf{P}_{EC}$ magnitude (left panels) and phase (right panels) as the probe tube was moved from the lateral EC wall (top) toward the umbo (bottom). Each $|\Delta\mathbf{P}_{EC}|$ curve shows one or more pronounced notches, and each $\angle\Delta\mathbf{P}_{EC}$ curve shows one or more 0.5-cycle steps. The frequencies of the first magnitude notch (black arrows in left panels) correspond closely (within 5%) to the frequencies midway in the first phase step (where $\angle\Delta\mathbf{P}_{EC}$ is 0.25 cycles—black arrows in right panels). The notch frequencies increased as the measurement location approached the umbo. (The curve for $\Delta\mathbf{P}_{EC}$ 1.0 mm from the umbo shows a double magnitude notch and a dip in $\angle\Delta\mathbf{P}_{EC}$ near 0.25 cycles. Sound pressure at this location was most variable among ears—discussed in Sec. V C.) A second magnitude notch and phase step were observed in the two measurements most distant from the umbo. In these measurements, the frequencies of the second magnitude notch (gray arrows in top left panels) correspond to the frequencies midway in the second phase steps (where $\angle\Delta\mathbf{P}_{EC}$ is 0.75 cycles—gray arrows in top right panels). These notches and phase steps indicate that the sound field becomes more complex at high frequencies but also indicate that the field has an ordered structure.

It will be convenient for comparisons of these results acquired at the MEEI to results acquired at Columbia to condense the curves of Fig. 5(b) into a single pair of panels (magnitude and phase). Figure 5(c) superposes the $\Delta\mathbf{P}_{EC}$ magnitude (top) and phase data (bottom) from Fig. 5(b). The changes in the $\Delta\mathbf{P}_{EC}$ magnitude notch and phase step frequencies with measurement location can be seen clearly. At frequencies where $|\Delta\mathbf{P}_{EC}|$ notches occur, the difference between the maximum and minimum $|\Delta\mathbf{P}_{EC}|$ among the measurement locations increases as the notch frequency increases.

Figure 6 shows data taken similarly at Columbia with the miniature fiber-optic microphone in the bony ear canal of animal ESO3 along a similar trajectory L [Fig. 2(b)], using the same sound coupler at the bony EC entrance. The results are very similar to those of Fig. 5. The sound pressure was measured over a wider spatial range than the measurements in Fig. 5 (to 4 mm from the umbo as opposed to 3 mm),[6] and the magnitude notch and phase change at the measurement locations furthest from the umbo occurred at a lower fre-
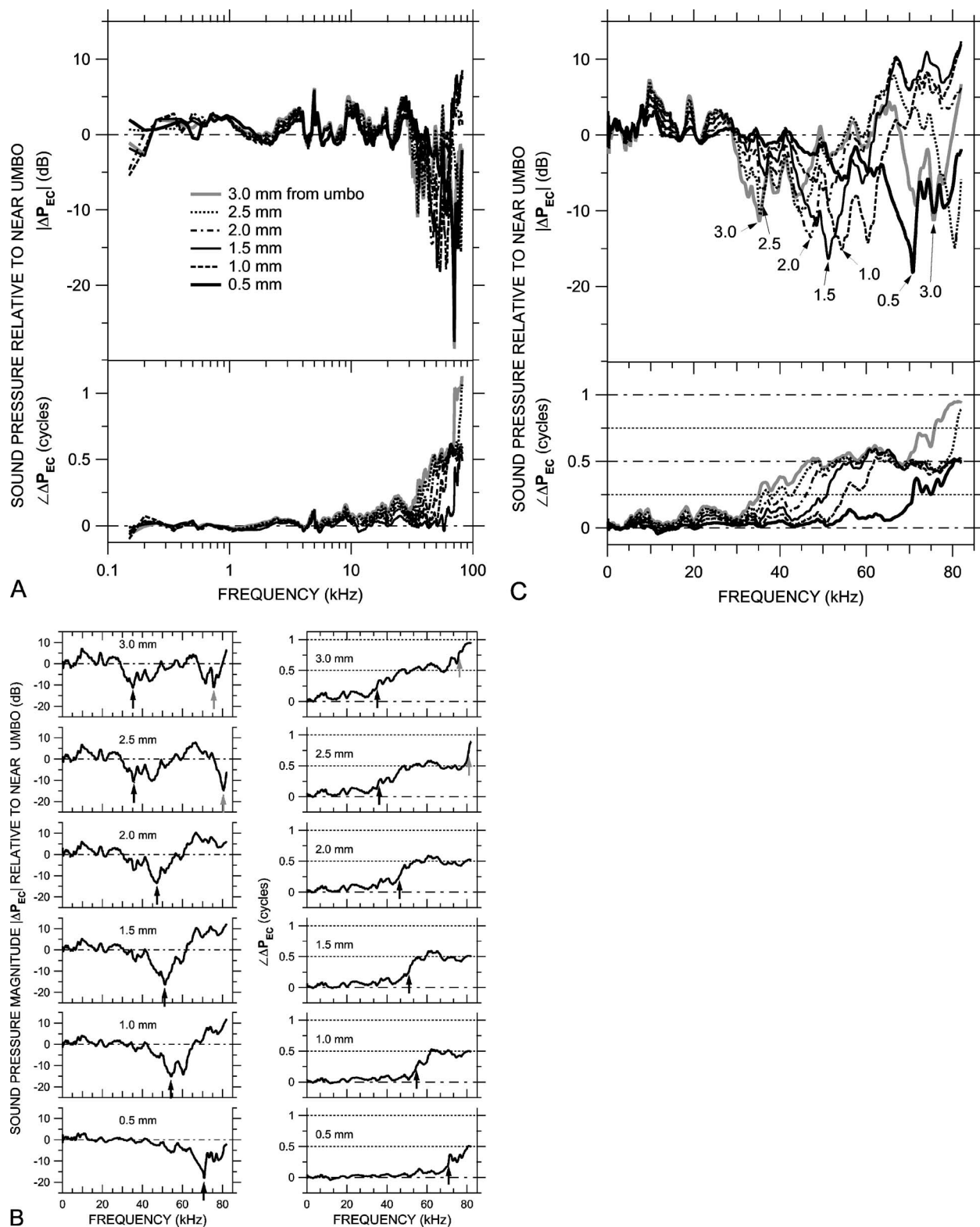
FIG. 5. Variations in ear canal sound pressure $\mathbf{P}_{EC}$ measured in 0.5-mm steps along trajectory L [see Fig. 2(b)] in ear 0402R with the probe tube microphone (MEEI) in response to a chirp stimulus. Shown is $\Delta\mathbf{P}_{EC}$, the ratio of $\mathbf{P}_{EC}$ in various locations along the trajectory to $\mathbf{P}_{EC}$ near the umbo. (A) $\Delta\mathbf{P}_{EC}$ at six locations, plotted on a logarithmic frequency scale. Variations in $\Delta\mathbf{P}_{EC}$ are small below 30 kHz. Top: Magnitude ratio in decibels; bottom: phase difference. (B) $\Delta\mathbf{P}_{EC}$ in the six locations, shown individually on a linear frequency scale. Left: Magnitude ratios in decibels; right: phase differences. Arrows indicate frequencies of $|\Delta\mathbf{P}_{EC}|$ notches and frequencies where $\angle\Delta\mathbf{P}_{EC}$ crosses 0.25 or 0.75 cycles. (C) The $\Delta\mathbf{P}_{EC}$ curves from (B) superposed, plotted on a linear frequency scale. Top: Magnitude ratio in decibels; bottom: phase difference.

quency (30 Hz vs 35 kHz in 0402R). There are many similarities between the two data sets: Both sets are clearly separable into low- and high-frequency regions, with the low-frequency regions (<25 kHz in Fig. 6) characterized by little

variation in magnitude (<6 dB) or phase (<0.1 cycle) and the high-frequency region characterized by spatial variations of 10–20 dB. The $|\Delta\mathbf{P}_{EC}|$ minima occur at approximately the same frequencies where $\angle\Delta\mathbf{P}_{EC}$ crosses 0.25 or 0.75 cycles
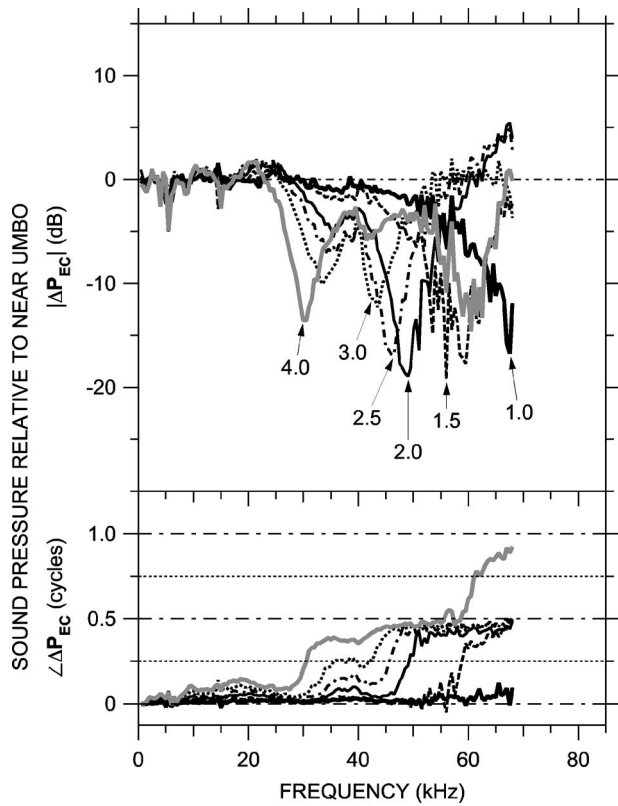
FIG. 6. Variations in ear canal sound pressure $\mathbf{P}_{EC}$ measured in 0.5-mm steps along trajectory L [see Fig. 2(b)] in ear ESO3 just post-mortem with the miniature fiber optic microphone (Columbia) in response to a series of tones. Shown is $\Delta\mathbf{P}_{EC}$ at six locations, computed as in Fig. 5 and plotted on a linear frequency scale. Top: Magnitude ratio in decibels; bottom: phase difference. Variations in $\Delta\mathbf{P}_{EC}$ with position and frequency are similar to those in Fig. 5.

($\pm 15\%$). In Fig. 6, as in Fig. 5, the difference between the maximum and minimum $|\Delta\mathbf{P}_{EC}|$ tended to increase above about 45 kHz. The noise floor in the fiber optic microphone was sufficiently high that $|\Delta\mathbf{P}_{EC}|$ at the notches should be considered as an upper bound at frequencies above 40 kHz, where the sound source output and thus the microphone signal-to-noise ratio were lower.

These same patterns (high-frequency $\Delta\mathbf{P}_{EC}$ magnitude notches and phase steps that increase in frequency as the measurement position approaches the umbo) were observed in three other ears at MEEI in which $\mathbf{P}_{EC}$ was measured near the umbo. Figure 7 shows $\Delta\mathbf{P}_{EC}$ computed from measurements made at the bony EC entrance and near the umbo from five animals—four at the MEEI (including 0402R, the animal in Fig. 5) and ESO3 (from Fig. 6). In these five ears the pressure transformation from the EC entrance to near the umbo was similar: $|\Delta\mathbf{P}_{EC}|$ in all ears had a minimum of $12-15$ dB and an associated phase transition near 35 kHz, and $|\Delta\mathbf{P}_{EC}|$ in three ears had a $6-10$ dB maximum near $60-65$ kHz. (The 3-mm distance between umbo and EC entrance is approximately $\frac{1}{4}$ wavelength at 30 kHz and $\frac{1}{2}$ wavelength at 60 kHz.) Figure 7 shows that the bony ear canal produces a pressure gain from EC entrance to umbo of $6-15$ dB between 30 and 50 kHz. The similarity between MEEI and Columbia data indicates that the larger probe tube used for the MEEI measurements had little effect on the EC sound field.
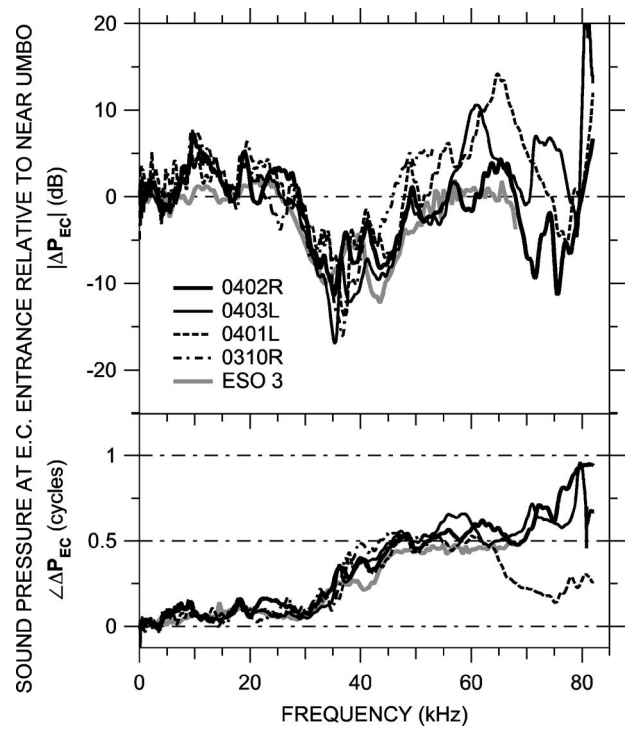
FIG. 7. $\Delta\mathbf{P}_{EC}$ near the opening of the bony ear canal in five ears measured at MEEI and Columbia. Top: Magnitude ratio in decibels; bottom: phase difference.

B. Sound field medio-axially and transversely across the ear canal

The sound field was also measured along two other trajectories across the ear canal. $\mathbf{P}_{EC}$ was measured in two ani-
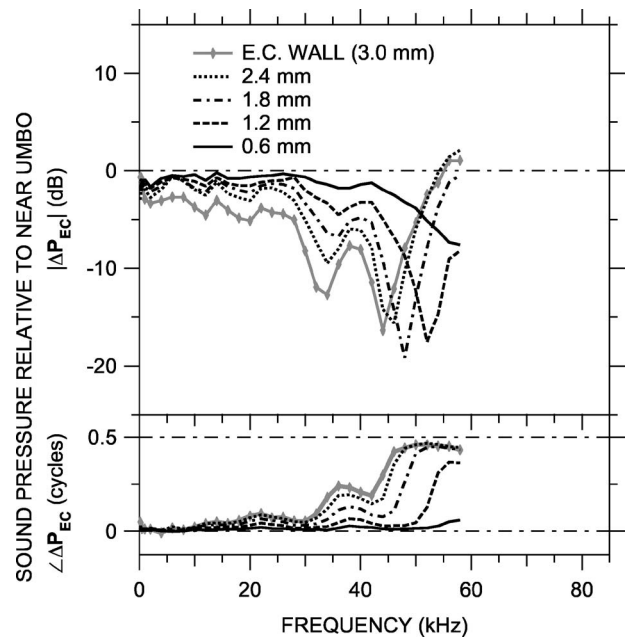
FIG. 8. Variations in ear canal sound pressure $\mathbf{P}_{EC}$ measured in 0.6-mm steps along a medio-axial trajectory "MA" [see Fig. 2(b)] in ear ESO2 with the miniature fiber optic microphone (Princeton) in response to a series of tones. Shown is $\Delta\mathbf{P}_{EC}$ at five locations, computed as in Fig. 5 and plotted on a linear frequency scale. Top: Magnitude ratio in decibels; bottom: phase difference. Variations in $\Delta\mathbf{P}_{EC}$ with position and frequency resemble those in Figs. 5 and 6 at locations 1.5 and 2.0 mm from the umbo.
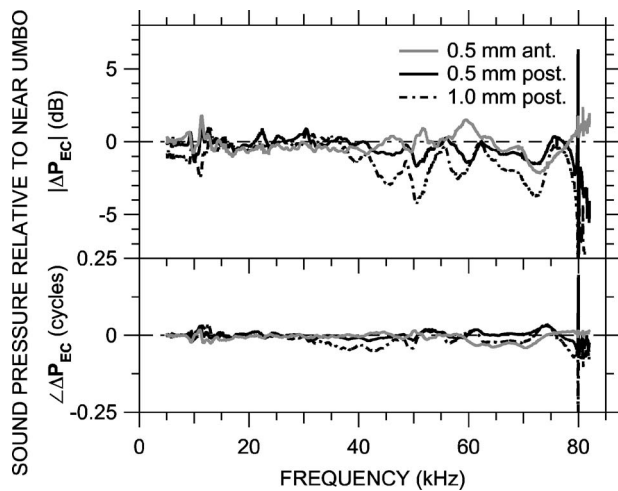
FIG. 9. Variations in ear canal sound pressure $\mathbf{P}_{EC}$ measured in 0.5-mm steps along an transverse trajectory 0.5 mm from the tympanic ring [see Fig. 2(c)] in ear 0305L (just post-mortem) with the probe-tube microphone (MEEI) in response to a chirp stimulus. Shown is $\Delta\mathbf{P}_{EC}$ at three locations, computed as in Fig. 5 and plotted on a linear frequency scale. Variations in $\Delta\mathbf{P}_{EC}$ are small below 70 kHz. Top: Magnitude ratio in decibels; bottom: phase difference.

mals at Princeton along a medio-axial trajectory "MA:" The sensor was placed in the ear canal through a small hole in the bony lateral wall and was advanced until it touched the umbo—see Fig. 2(b). This trajectory was more perpendicular to the tympanic ring than the approximately longitudinal trajectory L measured in the previous section. The results, shown for one animal as $\Delta\mathbf{P}_{EC}$ (computed as earlier) in Fig. 8, show similarities to those obtained from trajectory L (Figs. 5 and 6): There was <6 dB spatial variation in $\mathbf{P}_{EC}$ at frequencies below 30 kHz, and local minima in $|\Delta\mathbf{P}_{EC}|$ of up to 19 dB appeared at frequencies between 43 and 53 kHz. The minima were accompanied by a half-cycle increase in $\angle\Delta\mathbf{P}_{EC}$. (The upper frequency limit of these measurements was 58 kHz.) The $\Delta\mathbf{P}_{EC}$ curves in Fig. 8 generally resemble those in the third and fourth rows of Fig. 5(b), measured 2.0 and 1.5 mm from the umbo, respectively. Figure 2(b) shows that the trajectory MA samples the same section of the ear canal as the medial extent of trajectory L, so the similarity in $\Delta\mathbf{P}_{EC}$ measured along these two trajectories is not surprising.

Ear canal sound pressure was measured in three postmortem animals (due to anatomical constraints in the live ear—see Sec. III) at MEEI along the transverse trajectory shown in Fig. 2(c). The low height of the TM over the ear canal limited measurements to ±1 mm from the umbo. The results, shown normalized by $\mathbf{P}_U$ as in earlier figures for one animal in Fig. 9, are very different from $\Delta\mathbf{P}_{EC}$ along the longitudinal and medio-axial trajectories. Variations with position were less than 2 dB in magnitude and 0.05 cycles below 70 kHz, except slightly larger (~4 dB) near the EC wall above 40 kHz. Between 40 and 70 kHz the variations 0.5 mm away from the umbo are symmetric: for $|\Delta\mathbf{P}_{EC}|$ in decibels and $\angle\Delta\mathbf{P}_{EC}$ in cycles, $\Delta\mathbf{P}_{EC}$ at +0.5 mm $\approx -\Delta\mathbf{P}_{EC}$ at -0.5 mm. Similar results were obtained in two other animals. These results suggest that the largest variations in the sound

field within the terminal portion of the bony ear canal occur along the ear canal axis—variations in the transverse direction due to higher-order modes are small.

C. Stapes velocity with probe tube in various places in ear canal

The similarity of EC sound field measurements with the 1.27-mm-diam probe tube and the 0.17-mm-diam fiber optic microphone (Figs. 5–7) supports the idea that the probe tube did not disturb the sound field substantially. Another estimate of the perturbation of the EC sound field is the degree of variation in measured stapes velocity when the probe tube was advanced along the longitudinal trajectory within the ear canal. In the same ear in which the longitudinal $\Delta\mathbf{P}_{EC}$ was shown (g0402R—Fig. 5), stapes velocity was virtually unaffected by the position of the probe tube below 60 kHz: Variations were less than a factor of 2.

The above-described example was the "best case;" in the other MEEI ears shown in Fig. 7, the results were similar but the effects on stapes velocity occurred at lower frequencies (as low as 50 kHz). Our interpretation is that variations in probe-tube position within the ear canal caused the effect of the probe tube on the sound field to be larger or extend to lower frequencies in some ears compared to others. Another explanation of some of this variability in stapes velocity may be complexity in stapes motion at high frequencies (Olson and Cooper, 2000; Decraemer et al., 2007). Our results show that it is possible to measure $\mathbf{P}_{EC}$ at frequencies up to 60 kHz with a probe tube that is relatively large compared to the ear canal cross-section area (~30%) with little perturbation of the EC sound field near the umbo.

D. Sound field in an artificial ear

As a further test of the effects of the probe tube on the sound field in the ear canal, as well as a comparison of probe-tube measurements to other methods of estimating sound pressure at the TM, the sound pressure distribution was measured with the probe tube (at MEEI) in a simple artificial ear: the calibration coupler in the arrangement shown in Fig. 3(b). The earphone, sound delivery tube, and stimulus type (chirp) used were the same as for the EC measurements of Figs. 5, 7, and 9.

Two sets of measurements were made: In the first set, the flexible portion of the sound delivery tube was bent so the same probe tube used for $\mathbf{P}_{EC}$ measurements could be introduced through a hole in the wall and advanced axially toward the terminating microphone [Fig. 3(b)]. In this way, the sound pressure field $\mathbf{P}_{AE}$ within the artificial ear could be sampled in much the same way as $\mathbf{P}_{EC}$ was measured within the ear canal. In this setup, $\mathbf{P}_{AE}$ was measured at finer spatial increments (0.25 mm) than in the ear canal; representative measurements at 0.5-mm spacing normalized by $\mathbf{P}_{AE}$ near the terminating microphone ($\Delta\mathbf{P}_{AE}$) are shown in Fig. 10(a). A comparison with $\Delta\mathbf{P}_{EC}$ in Fig. 5(b) shows a similar ordered structure for measurements 2.0 mm or more from the termination: each such $|\Delta\mathbf{P}_{AE}|$ curve (left panels) shows one or more pronounced notches, and each such $\angle\Delta\mathbf{P}_{AE}$ curve (right panels) shows one or more 0.5-cycle steps. Frequen-

cies of the first magnitude notch (black arrows in left panels) correspond closely (within 2%) to the frequencies midway in the first phase steps (where $\angle \Delta \mathbf{P}_{AE}$ is 0.25 cycles—black arrows in right panels). The notch frequencies increased as the distance from the probe tube to the termination decreased. A second notch and corresponding phase shift were observed in the more distal measurements (gray arrows). The depth of the first $|\Delta \mathbf{P}_{AE}|$ notches (measured maximum-to-minimum) was about 27 dB and roughly constant across frequency. Figure 10(b) shows the curves superposed as in Fig. 5(c).

During these first measurements, the output of the $\frac{1}{8}$-in. reference microphone was recorded as well to estimate the effect of the probe tube on the sound field in the artificial ear and, by extension, in the ear canal. Changes in reference microphone output were generally about 3 dB below 60 kHz. Between 60 and 80 kHz variations were slightly larger (5 dB) and showed a decrease in $|\mathbf{P}_{AE}|$ consistent with shadowing. The magnitude of these effects was substantially less than the longitudinal $\mathbf{P}_{AE}$ variations, which suggests that the probe tube did not disrupt the sound field substantially.

In the second set of measurements, the probe tube microphone [Fig. 3(a)] was replaced with a plug, and the brass portion of the sound delivery tube was withdrawn approximately 3 mm from the calibration coupler. The dimensions of this extra space approximated the gerbil ear canal dimensions for a "uniform-tube calibration" (Pearce *et al.*, 2001) or "artificial-cavity calibration" (Overstreet and Ruggero, 2002). The sound pressure at the termination measured with the reference microphone is compared with $\mathbf{P}_U$ in Sec. V F.

V. DISCUSSION

A. Basic description of EC sound field from measurements *in situ*

Our *in situ* measurements of sound pressure $\mathbf{P}_{EC}$ at multiple points within the bony gerbil ear canal (Figs. 5–9) lead to the following description of how sound pressure varies in the ear canal near the tympanic membrane:

(1) The measurements along approximately longitudinal trajectories indicate that at frequencies below 30 kHz the spatial dependence of the measured sound pressure is small [Figs. 5(a) and 7]. $\mathbf{P}_{EC}$ measured at any frequency in this range is relatively independent of axial position: Spatial variations are less than $\sim$5 dB in magnitude and 0.1 cycles in phase. At frequencies above 30 kHz, spatially dependent variations in $\mathbf{P}_{EC}$ were observed that were $>$15 dB in magnitude and 0.25–1 cycle in phase. Therefore, measurements of the sound pressure at a single point within the bony ear canal provide a reasonably accurate estimate of sound pressure at the TM at frequencies less than 30 kHz, but above 30 kHz the accuracy of a remote TM sound pressure estimate diminishes (see the discussion in Dong and Olson, 2006).

(2) The sound pressures measured along our longitudinal trajectories show pronounced minima in the frequency domain that coincide with half-cycle phase transitions (Figs. 5 and 6). The frequencies of the minima and phase transitions vary with axial position: Minima occurred at
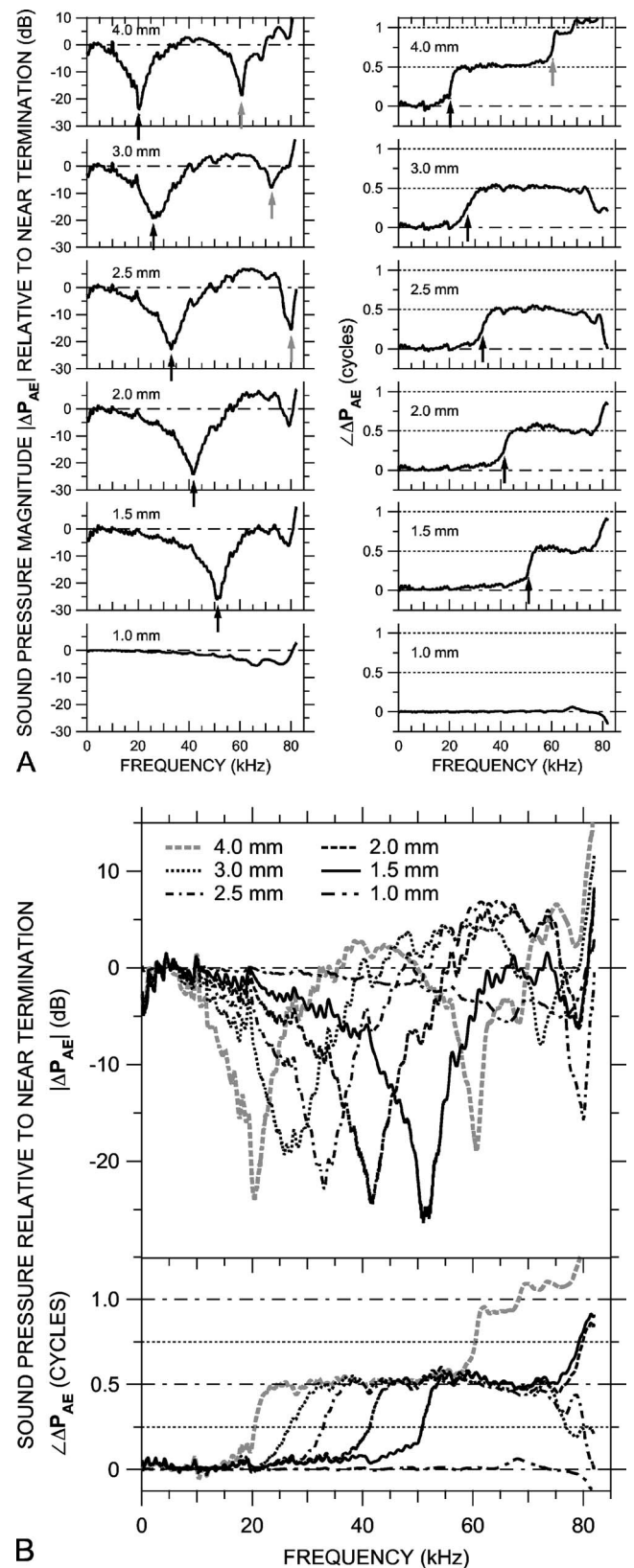
FIG. 10. Variations in sound pressure $\mathbf{P}_{AE}$ measured axially in 0.5-mm steps within an artificial ear with the probe-tube microphone (at MEEI) in response to a chirp stimulus. The artificial ear was terminated with a $\frac{1}{8}$-in. condenser microphone. (A) $\Delta \mathbf{P}_{AE}$, the ratio of $\mathbf{P}_{AE}$ at six locations to $\mathbf{P}_{AE}$ near the termination, shown individually on a linear frequency scale. Left: Magnitude ratios in decibels; right: phase differences. Arrows indicate frequencies of $|\Delta \mathbf{P}_{AE}|$ notches and frequencies where $\angle \Delta \mathbf{P}_{AE}$ crosses 0.25 or 0.75 cycles. (B) The $\Delta \mathbf{P}_{AE}$ curves from (A) superposed, plotted on a linear frequency scale. Top: Magnitude ratio in decibels; bottom: phase difference.

30–35 kHz in measurements 2.5–3.0 mm from the umbo (near the opening in the lateral EC wall) and at 55–60 kHz in measurements within 1 mm of the umbo. $\mathbf{P}_{EC}$ along a medio-axial trajectory (Fig. 8) had a frequency dependence that was similar but compressed in frequency. As will be discussed in Sec. C, the frequency and location dependence of these minima and their associated phase changes suggest the presence of standing waves in the terminal portion of the ear canal.

(3) Sound pressures measured along a transverse trajectory in the plane of the tympanic ring about 0.5 mm from the umbo are generally uniform, varying by less than 4 dB in magnitude at frequencies up to ~70 kHz (Fig. 9). This result suggests that below 70 kHz, the concept of a stimulus sound pressure "at" the gerbil TM is reasonable. This result also suggests that the sound field in other species with simpler ear canals is likely to be simple near the TM.

B. Effects of measurement conditions on ear canal sound field

The conclusions drawn about the nature of the ear canal sound field rely on an assumption that any effect the probe tube has on the sound field is small: Neither occlusion of the ear canal nor any sound shunt path provided by the probe tube is significant. The probe tube occluded the EC cross-section area by about 30%. Several lines of evidence suggest that the effect of the probe tube on the sound field was minor: (1) In general, stapes velocity changed by less than a factor of 2 below about 60 kHz as the probe tube was moved into and out of the ear canal. (2) When the probe tube was introduced into the calibration coupler or advanced in the artificial ear (as for Fig. 10), the output of the terminating microphone changed by less than 3 dB. (3) The diameter of the miniature microphone (used at Princeton and Columbia) was only 15% of that of the probe tube (used at MEEI), yet longitudinal sound pressure variations measured with the miniature microphone (Fig. 6) were nearly identical to those measured with the probe tube (Fig. 5). (4) The spatial locations of $|\mathbf{P}_{AE}|$ notches in the artificial ear match their theoretical location, and the depth of $|\mathbf{P}_{AE}|$ notches was independent of probe tube location (see Sec. V C). These lines of evidence suggest that the presence of the probe tube did not change the nature of the sound field in the ear canal.

Though the measurements at MEEI along a longitudinal trajectory were made in a live preparation, the longitudinal measurements at Columbia were made post-mortem, as were measurements at MEEI along a transverse trajectory. Past studies have shown little effect of death on middle-ear input or transmission properties: For instance, Rosowski *et al.* (1990) demonstrated that the middle-ear input admittance of guinea pigs was not greatly affected by the death of the animal. While post-mortem changes in active cochlear mechanics are large, passive intracochlear pressure and BM motion change little post-mortem as long as the cochlear fluid is not depleted (Olson, 2001; Ren and Nuttall, 2001; Cooper, 2000), so there should be little effect on middle-ear input impedance or $\mathbf{P}_{EC}$. Measurements of longitudinal sound pressure patterns in one ear at MEEI post-mortem versus alive

showed only small differences except above 50 kHz: $|\mathbf{P}_{EC}|$ notch depths were within 1 to 2 dB of those in live ears below 50 kHz and 5 dB or so less at higher frequencies.

C. Evidence for longitudinal standing waves within the ear canal

The sound pressure spectra measured at various locations along an approximately longitudinal trajectory in the ear canal (Figs. 5 and 6) have much in common with the theoretical frequency dependencies associated with a standing wave pattern [Fig. 1(b)]. In each case, spectra at locations far from the TM have two magnitude notches and a 0.5-cycle phase step at the notch frequencies, and the notch/phase step frequencies increase as the observation location is moved toward the TM. A comparison of Fig. 5(c) with Fig. 1(b) suggests that the sound field in the ear canal is a combination of a traveling wave and a standing wave, with $B/A \approx 0.7$ near 30 kHz.

The ear-canal and artificial-ear spectra of Figs. 5, 6, and 10 also have some features that are not present in the theoretical curves of Fig. 1(b): Fig. 5(c) shows double $|\Delta\mathbf{P}_{EC}|$ notches and $\angle\Delta\mathbf{P}_{EC}$ ripple at 1.0 mm from the umbo (and to a lesser extent, 2.5 mm), and Figs. 6 and 10 show $|\Delta\mathbf{P}_{EC}|$ >0 dB and $\angle\Delta\mathbf{P}_{EC}$<0.5 cycles at higher frequencies. These differences suggest the presence of higher-order modes at high frequencies, possibly generated by the TM or by changes in ear canal cross section. We will discuss the limitations of higher-order modes on reflectance estimates in the next section.

Another more descriptive demonstration of the similarity between our measurements and the theoretical standing wave patterns of Fig. 1 is provided by constructing the spatial dependence of the sound pressure field from spectra measured at various locations within the ear canal or the artificial ear. Each curve in Fig. 11(a) shows the spatial distribution of $\Delta\mathbf{P}_{EC}$ at a frequency where one of the $\Delta\mathbf{P}_{EC}$ spectra in Fig. 5(c) had a magnitude minimum. Each such isofrequency curve was constructed from the data in Fig. 5(c) in the following manner: (1) Frequencies at which $|\Delta\mathbf{P}_{EC}|$ had a clear minimum at any measurement position were identified; (2) $|\Delta\mathbf{P}_{EC}|$ at these frequencies at each measurement position was extracted; (3) the maximum $|\Delta\mathbf{P}_{EC}|$ among the measurement positions at each of the frequencies was identified and used to normalize the other $|\mathbf{P}_{EC}|$ measurements at that frequency; and (4) the normalized $|\Delta\mathbf{P}_{EC}|$ values at each frequency were plotted as a function of measurement position in the top panel of Fig. 11(a). A similar procedure was used to prepare the plots of $\angle\Delta\mathbf{P}_{EC}$ versus position in the bottom panel of Fig. 11(a), though the minimum $\angle\mathbf{P}_{EC}$ was used as the normalization factor. Figure 11(b) shows similar curves at the frequencies of the magnitude minima in the $\Delta\mathbf{P}_{AE}$ spectra in Fig. 10. (Note that these frequencies are not intrinsically special; if different axial positions had been sampled, the minima would have occurred at different frequencies.) The data in Figs. 5 and 10 were finely sampled in frequency (50-Hz spacing within the chirp), so it was possible to choose a frequency where a pronounced dip was present and be sure that the depth of the minimum was captured.
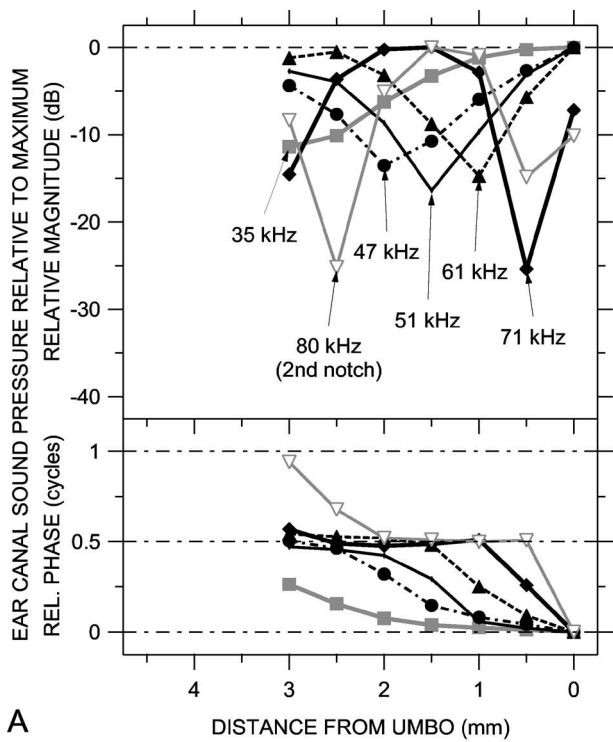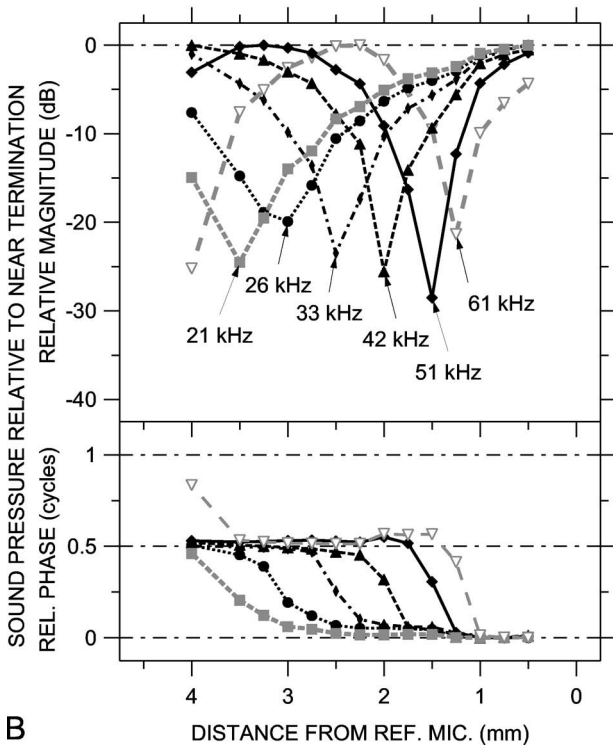
FIG. 11. Standing wave patterns at frequencies of notches in sound pressure spectra, constructed from sound pressure spectra measured at various locations. Because the MEEI data were taken with a chirp (broadband) stimulus, the frequencies at which notch minimums occurred could be identified reliably. (A) Standing wave patterns in ear 0402R computed from $\Delta\mathbf{P}_{EC}$ at various EC locations, from Fig. 5(c). (B) Standing wave patterns in the artificial ear computed from $\Delta\mathbf{P}_{AE}$ at various locations, from Fig. 10(b). Top: Magnitude at frequencies of magnitude notches; bottom: Phase at frequencies of magnitude notches.

The spatial sound pressure distributions in Fig. 11 show that the sound fields in the ear canal and the artificial ear have the same basic structure as the theoretical standing
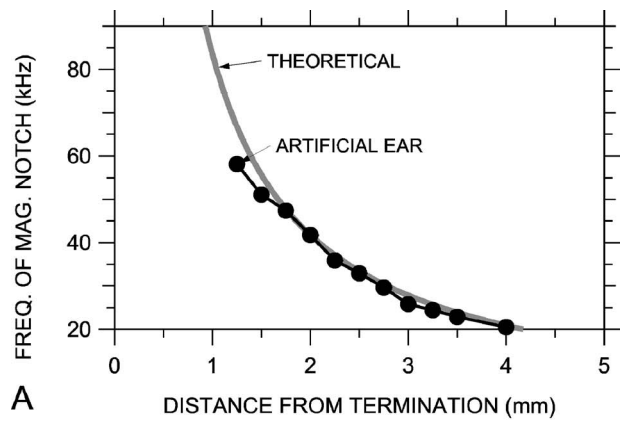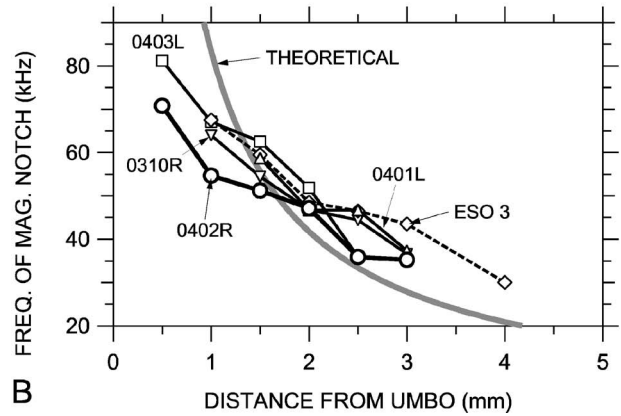
FIG. 12. Comparison of the frequencies of sound pressure magnitude notches at measurement locations to the theoretical $\frac{1}{4}$-wave notch frequencies and locations in a rigidly terminated uniform tube (thick gray line). (A) In the artificial ear terminated by a reference microphone, 0.25 mm spacing—closed symbols. (B) In five ear canals.

wave patterns of Fig. 1(a). Specifically, the spatial distributions of both $\Delta\mathbf{P}_{EC}$ in Fig. 11(a) and $\Delta\mathbf{P}_{AE}$ in Fig. 11(b) show a pronounced magnitude notch and a half-cycle phase accumulation from one side of the notch to the other. At the highest frequencies an additional magnitude notch and half-cycle phase accumulation can be seen at the measurement point most distant from the umbo or termination. In the ear canal, the notches at the high frequencies of 71 and 80 kHz are deeper than the notches at lower frequencies (35–61 kHz); in the artificial ear, the depth of the notches is roughly constant. A comparison of Fig. 11(a) with Fig. 1(a) suggests that the magnitude of the pressure reflection coefficient B/A in the ear canal is approximately 0.9 near 70 kHz and closer to 0.7 at lower frequencies. The similarities of both the EC sound pressure spectra and spatial distributions to the theoretical spectra and standing wave patterns of Fig. 1 support the notion that the sound field in the ear canal is a combination of a traveling wave and a standing wave.

A more quantitative comparison of the salient features of $\Delta\mathbf{P}_{EC}$ and EC standing wave patterns to the theoretical sound field is shown in Fig. 12. In Fig. 12, the frequencies of the first (lowest-frequency) $|\Delta\mathbf{P}_{EC}|$ and $|\Delta\mathbf{P}_{AE}|$ notches at the various measurement locations are compared to the theoretical location of the first sound pressure null in the artificial ear. We assume that the reference microphone is a rigid termination (see Sec. V D), so the distance from the null to the

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Ravicz *et al.*: Gerbil ear canal sound pressure distribution   2167
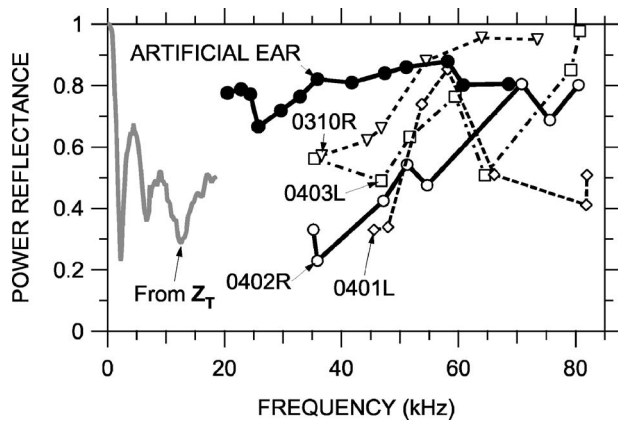
FIG. 13. Power reflectance $\mathfrak{R}$ computed from standing wave patterns in the artificial ear and in several gerbil ear canals. Shown are: $\mathfrak{R}_{AE}$ in the artificial ear (thick solid line and closed symbols), computed from standing wave patterns of Fig. 11(b); $\mathfrak{R}_{EC}$ in ear 0402R (thick solid line, open symbols), computed from standing wave patterns of Fig. 11(a); $\mathfrak{R}_{EC}$ in other MEEI ears (thin dashed lines and open symbols), computed in a similar fashion; and $\mathfrak{R}$ computed from the mean middle-ear input impedance $\mathbf{Z}_T$ between 100 Hz and 18 kHz in 5 gerbil ears (gray line, from Ravicz *et al.*, 1996).

termination is equal to $\lambda/4$ (e.g., Kinsler *et al.*, 1982). The locus of $|\Delta\mathbf{P}_{AE}|$ notch frequencies in the artificial ear [Fig. 12(a)—closed symbols] matches the theoretical curve quite well in location and shape.

Frequencies of $|\Delta\mathbf{P}_{EC}|$ notches [Fig. 12(b)—open symbols] vary somewhat from the theoretical: At locations fairly remote from the umbo, where the first notch occurs below about 40 kHz, the slope of the notch frequency versus distance curve is similar to the theoretical. The curves are offset to the right, which implies that the effective length of the ear canal was less than its actual length. At locations closer to the umbo, where the notch occurs between 45 and 55 kHz, the slope is shallower, as if the effective length of the ear canal were increasing. Possible reasons for this apparent stretching of the standing-wave pattern include the following (1) Variations in the ear canal cross-section area. These variations affect the acoustics in ways that can be interpreted as changing the effective EC length (Benade, 1990, pp. 473–476). (2) A change in the nature of the TM impedance. If the TM impedance became more compliance-dominated or less mass-dominated as frequency increased, a phase shift in the sound wave reflected from the TM could make the ear canal appear longer.[7] (3) A mismatch between the trajectory of the probe tube and the effective axis of the curved ear canal. If the probe-tube trajectory is along the outside of the curve (so that normal planes evenly spaced along the axis diverge), the spacing between magnitude notches could be larger than along the EC axis (e.g., Stinson, 1990). (4) Effects of the EC flare and curvature near the TM that limit the utility of a description of the ear canal as a (nearly) uniform tube with a normal termination. In cat and human, the TM terminates the ear canal obliquely so that the TM may be considered a compliant section of EC wall, and this oblique termination increases the complexity of the standing wave pattern (Stinson and Khanna, 1989; Stinson, 1990). The EC geometry in gerbil is even more complicated than in these two species, so

a simple one-dimensional model may not be adequate to predict $|\mathbf{P}_{EC}|$ notch locations to a high degree of accuracy; a 3D model may be necessary.

D. Computation of middle-ear power reflectance and absorption from standing wave patterns

The standing wave ratios from standing wave patterns of Fig. 11(a) and spectra of Fig. 5 are used in Eq. (3) to compute the ear-canal sound power reflectance $\mathfrak{R}_{EC}$, the fraction of the incident sound power that is reflected at the tympanic membrane (and not transmitted into the middle ear). We also compute the reflectance $\mathfrak{R}_{AE}$ from SWRs in the artificial ear [Fig. 11(b)], and we compute the theoretical reflectance $\mathfrak{R}_{theo}$ of the calibration coupler from its characteristic impedance $Z_0 = 4\rho_0 c/\pi d^2$ (from Sec. II) and the impedance of the terminating microphone $\mathbf{Z}_{mic}$ by

$$\mathfrak{R}_{theo} = \left(\frac{\mathbf{Z}_{mic} - Z_0}{\mathbf{Z}_{mic} + Z_0}\right)^2 \qquad (4)$$

(e.g., Kinsler *et al.*, 1982).[8] Discrepancies between the measured and theoretical reflectances are used to estimate errors and limits of ear-canal reflectance measurements.

Figure 13 shows that $\mathfrak{R}_{AE}$ computed from standing wave patterns in the artificial ear is about 0.7–0.8 [from Fig. 11(b)—thick line and closed circles]. This value is somewhat lower than the theoretical reflectance $\mathfrak{R}_{theo} \approx 1$ across the frequency range. Because $\mathfrak{R}_{AE}$ was computed from the depth of the $|\Delta\mathbf{P}_{AE}|$ notches, this discrepancy implies that real or apparent sound pressure not associated with the incident or reflected plane wave was present at the notch frequencies, thereby making $|\Delta\mathbf{P}_{AE}|$ at the notch greater than it would be were this other sound pressure not present. A closer look at Fig. 10(b) shows that some of the $|\Delta\mathbf{P}_{AE}|$ spectral notches appear flattened at the bottom, in contrast to the sharp tips in the theoretical spectra in Fig. 1(b). The presence of extra sound pressure at the $|\Delta\mathbf{P}_{AE}|$ notches would cause the SWR and hence $\mathfrak{R}_{AE}$ to be underestimated and puts an upper bound of about 0.8 on the reflectance measurable by this method.

Possible mechanisms for notch-flattening include: apparent sound pressure from environmental noise, microphone self-noise, or probe tube artifact (see Sec. III D); and real sound pressure from higher-order modes. We examine these possibilities in turn: (1) The noise floor (includes environmental and microphone noise) was 5–10 dB below the probe tube artifact, which was itself at least 30 dB below the microphone signal in the calibration (see Sec. III). If the SWR is limited by artifact to a maximum of 30 dB (ratio of $|\mathbf{P}_{AE}|$ at notch to $|\mathbf{P}_{AE}|$ at maximum=0.03), the maximum measurable $\mathfrak{R}_{AE}$ [by Eq. (3)] is 0.89. This result suggests that there is some other source of the extra sound pressure. (2) Higher-order modes may be generated at the termination and decay over a small but finite distance that increases at frequencies close to the propagation frequency (88 kHz for the 2.3-mm calibration coupler). These higher-order modes do not transmit energy below their propagation frequencies (Pierce, 1981, p. 316), but their presence complicates estimates of the standing wave ratio. A maximum measurable $\mathfrak{R}_{AE}$ of 0.8

Ravicz *et al.*: Gerbil ear canal sound pressure distribution

suggests [by rearranging Eq. (3)] that the sound pressure contributed by higher-order modes in the artificial ear is $\sim 0.1A$ (the sound pressure in the incident plane wave) or ~ 26 dB below the maximum $|\mathbf{P}_{AE}|$. The result that transverse $\mathbf{P}_{EC}$ variations near the TM are small (1 to 2 dB; see Fig. 9) implies that the contribution of higher-order modes to EC sound pressure is comparable to that in the artificial ear. This implies in turn that, though measurement of reflectances >0.8 are problematic, reflectances below 0.8 can be measured accurately.

Figure 13 also shows the power reflectance $\mathfrak{R}_{EC}$ computed from standing wave patterns in ear canals. $\mathfrak{R}_{EC}$ in animal 0402R [from Fig. 11(a)—thick line and open circles] was relatively low at the lower frequencies (0.2–0.3 near 35 kHz) and increased at higher frequencies to a value comparable to that measured in the artificial ear (0.7–0.8 near 70 kHz). Because 0.8 is the maximum reflectance we can measure, it is possible that $\mathfrak{R}_{EC}$ is even higher at high frequencies. Similar results were obtained in three other animals (thin lines and open symbols). The variability in $\mathfrak{R}_{EC}$ was about 0.1 among several measurement sets in animal 0402R and about 0.2 in the other ears. In all ears $\mathfrak{R}_{EC}$ was low at 35–45 kHz and increased with frequency.[9]

Figure 13 also includes the power reflectance $\mathfrak{R}_T$ computed by Eq. (4) from previous measurements of gerbil middle-ear input impedance $\mathbf{Z}_T$ at lower frequencies (Ravicz *et al.*, 1992, 1996—gray line). $\mathfrak{R}_T$ at the highest frequency computed by this measure (~ 0.5 at 18 kHz) was similar to $\mathfrak{R}_{EC}$ at the lowest frequencies computed from standing wave patterns (~ 30 kHz). $\mathfrak{R}_T$ increased to ~ 1 below 1 kHz. Figure 13 indicates that more incident sound power is transmitted into the gerbil middle ear (less incident sound power is reflected) in the middle frequencies than at very low (<100 Hz) or high frequencies (>50 kHz).

The result that $\mathfrak{R}_{EC}$ is comparable to $\mathfrak{R}_T$ near 30 kHz and increases at higher frequencies suggests that the gerbil $|\mathbf{Z}_T|$ increases with frequency. Because the reflectance is a measure of the impedance mismatch between the terminating impedance and the frequency-independent characteristic impedance Z_0 [Eq. (4)], a large reflectance implies a large difference between the magnitude of the terminating impedance and Z_0. Similarly, increasing $\mathbf{Z}_T$ artificially (for example, by replacing the TM with a nearly rigid microphone diaphragm) can be expected to cause an increase in reflectance and spatial $\mathbf{P}_{EC}$ variations at lower frequencies.

As mentioned earlier, the fraction of incident sound power that enters the middle ear is equal to $1-\mathfrak{R}_{EC}$. Over much of our measurement range $\mathfrak{R}_{EC}$ was 0.2–0.6, which means that 40%–80% of the power in the ear canal between 6 and 50 kHz entered the middle ear. At higher frequencies, $\mathfrak{R}_{EC} \geq 0.8$, which means that at most 20% of the EC power entered the middle ear. The increase in $\mathfrak{R}_{EC}$ means that less energy is available to be transmitted to the cochlea at higher frequencies (Rosowski *et al.*, 1986, 1988).

The detection of sound requires both transmission of a sound signal to the cochlea and activation of sensory cells within the cochlea (e.g., Dallos, 1996). Ruggero and Temchin (2002) point out that the high-frequency limit of hearing (the audiometric limit) corresponds to the upper ex-

tent of the cochlear sensitivity map. We have shown here that ear canal reflectance also increases as sound frequency approaches the audiometric limit, which indicates that sound transmission to the cochlea decreases. It is likely that the frequency ranges of sound transmission to the cochlea and the cochlear sensitivity map are similar so that the capabilities of one do not greatly surpass the limitations of the other.

E. Comparisons to ear canal sound pressure field and reflectance in other species

Our $\Delta \mathbf{P}_{EC}$ data are similar to the standing wave patterns measured previously in human (Stinson, 1985b; Lawton and Stinson, 1986; Stinson, 1990; Chan and Geisler, 1990) and cat (Khanna and Stinson, 1985; Stinson and Khanna, 1994) in a fashion similar to that described here, though not to nearly such high frequencies. The frequencies of standing-wave notches we observed in gerbil are higher than those observed in ear canals of these other species, consistent with the shorter remaining gerbil ear canal.

Our estimates of gerbil $\mathfrak{R}_{EC}$ in Fig. 13 show the same frequency dependence as $\mathfrak{R}$ computed previously in human, chinchilla, and cat (in healthy ears). The reflectance computed for human from $\mathbf{Z}_T$ by Eq. (4) (Voss and Allen, 1994; Margolis *et al.*, 1999) was ~ 1 at 250 Hz and below, decreased to 0.3–0.5 near 1 kHz, and increased again to about 0.7 from 7 to 10 kHz (the highest frequency of reliable measurements), while $\mathfrak{R}$ computed from SWR increased from about 0.25 near 4 kHz (the lowest frequency measured) to 0.7 near 8 kHz [and decreased slightly as frequency increased further to 13 kHz, the highest frequency measured; Stinson (1990)]. Chinchilla $\mathfrak{R}$ was similar to human except that $\mathfrak{R}$ was high below 2 kHz (Margolis *et al.*, 2001). Previous estimates of cat $\mathfrak{R}$ are more variable: Cat $\mathfrak{R}$ computed from $\mathbf{Z}_T$ (Lynch *et al.*, 1994; Puria and Allen, 1998) was ~ 1 below 300 Hz, decreased to ≤ 0.3 at 1.5 kHz, and remained nearly constant (except for a peak near 4 kHz in the Lynch *et al.* data) to the highest frequency measured (22 kHz, Lynch *et al.*; 32 kHz, Puria and Allen). Cat $\mathfrak{R}$ computed from SWR (Stinson and Khanna, 1994) was quite variable at high frequencies: Ears with TMs in poor condition had high $\mathfrak{R}$, but the others showed $\mathfrak{R}$ generally low (<0.3) at 10 kHz (the lowest frequency measured) and increasing sharply between 23 kHz and the highest frequency measured (generally 32 kHz), in contrast to $\mathfrak{R}$ computed from $\mathbf{Z}_T$ by the other authors. Our results suggest that ear canal reflectance in gerbil is representative of $\mathfrak{R}$ in other species: relatively low over a wide middle frequency range, with higher values at low and high frequencies.

F. Comparison with ear canal sound pressure estimated by other methods

Our data allow us to compare the accuracy of the three methods of estimating TM sound pressure described in Sec. I: shortening the ear canal and using a probe tube to measuring the sound close to the TM, mathematically modeling the ear canal, and using an artificial ear.

It should be noted that alternatives to sound pressure as a measure of middle-ear input have been proposed: For ex-

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Ravicz *et al.*: Gerbil ear canal sound pressure distribution   2169
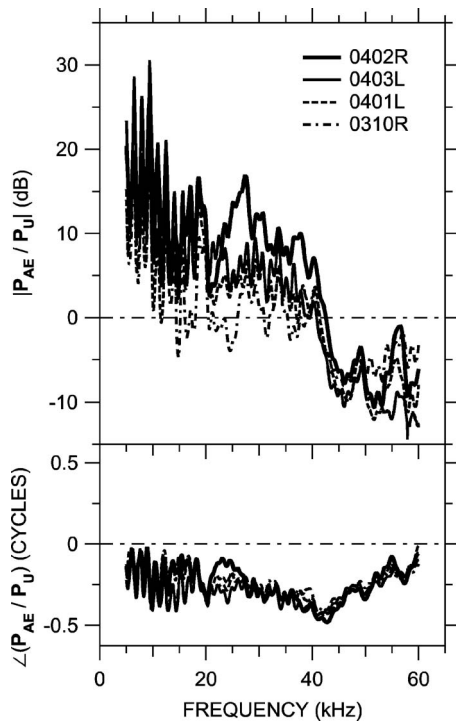
FIG. 14. Ratio of sound pressure $\mathbf{P}_{AE}$ measured at the terminating microphone of the artificial ear to $\mathbf{P}_U$ measured near the umbo in four gerbil ears. The length of the artificial ear was adjusted to mimic the length of the gerbil bony ear canal. Data are not shown below 5 kHz because leaks between the probe tube and the ear canal influence the results. Top: Magnitude; bottom: phase. The artificial ear calibration overestimates $\mathbf{P}_U$ at low frequencies and underestimates $\mathbf{P}_U$ between 40 and 60 kHz.

ample, ear-canal sound intensity (e.g., Neely and Gorga, 1998), in part because the spatial distribution of intensity in the ear canal, is much more uniform than the sound pressure field (Neely and Gorga, 1998). A disadvantage of using intensity as input is that the determination of intensity from sound pressure measurements requires knowledge of the TM impedance (e.g., Keefe *et al.*, 1993; Neely and Gorga, 1998), a quantity that may not be easily measurable at high frequencies (e.g., Stinson *et al.*, 1982; Khanna and Stinson, 1985). Also, analysis of middle-ear mechanics requires knowledge of stimulus phase; when intensity is computed from sound pressure, $\angle\mathbf{P}_{EC}$ information is lost. The use of $\mathbf{P}_U$ to specify the middle-ear input preserves the phase information present in the ear-canal stimulus. For these reasons as well as the relative simplicity of the measurement, ear canal sound pressure remains the most commonly used measure of middle-ear input.

We shortened the gerbil ear canal and measured the sound pressure near the ear canal entrance and umbo (the first method mentioned previously) and found that longitudinal and transverse $\mathbf{P}_{EC}$ variations were less than 6 dB at frequencies less than 30 kHz. This result implies that $\mathbf{P}_{EC}$ measured at the entrance of the bony ear canal is a fairly accurate and repeatable estimate of the sound pressure input to the middle ear at frequencies below 30 kHz, which includes much but not all of the gerbil range of hearing. The larger longitudinal $\mathbf{P}_{EC}$ variations at 40 kHz and above indicate that measurements of middle-ear function at these high frequen-

cies should measure the sound pressure near the umbo if possible, and if that is not practical, must be interpreted with the measurement location in mind.

The only other EC sound pressure distribution data in gerbil of which we are aware were measured by Maki and Furukawa (2005a, Fig. 2i) with a probe tube introduced down an intact ear canal post-mortem. For these measurements the TM was stiffened with cyanoacrylate adhesive. Notches were observed in the EC transfer function of apparently 16–23 dB at 14–32 kHz, deeper than the 12- to 15-dB notches we observed near 35 kHz. Though the location of their probe-tube tip relative to the TM is not known, the lower notch frequency suggests that it was lateral to the bony ear canal. The increased notch depth may be due to the stiffening of the TM, which probably increased the middle-ear input impedance and therefore EC reflectance and, as a result, increased $\mathbf{P}_{EC}$ variations.

The second method, mathematical modeling of the gerbil bony ear canal, is difficult due to the complex shape of the ear canal: If the EC axis is considered to terminate at the umbo, then the axis curves sharply near the TM; and if the EC axis is considered to extend to the medial edge of the tympanic ring, then the TM forms part of the EC wall (see Fig. 2). The ear canal also flares as it nears the TM. A simple model that neglects EC curvature and considers the TM to terminate the ear canal (Ravicz *et al.*, 1996) fit $\mathbf{Z}_T$ data fairly well up to 18 kHz, but its applicability at higher frequencies is unknown, and we did not pursue this model further. Stinson (1990) developed a model of the human ear canal that successfully predicted sound pressure at the TM from more lateral $\mathbf{P}_{EC}$ measurements and an accurate representation of the ear canal shape. The more complicated shape of the gerbil ear canal may render this approach more difficult.

Regarding the third method, use of an artificial ear, our data show that the location and range of $\mathbf{P}_{EC}$ variations are affected by the shape and termination of the ear canal. For much of the frequency region in which $|\mathbf{P}_{EC}|$ notches were apparent, they were shallower than those in the artificial ear. The result in this frequency region highlights the greater power reflectance of the artificial ear, a difference that is expected to produce a bias when a calibration performed in an artificial ear [a uniform-tube calibration (Pearce *et al.*, 2001) or artificial-cavity calibration (Overstreet and Ruggero, 2002)] is used to set sound stimulus levels in an actual ear (e.g., Siegel, 1994).

Figure 14 shows the ratio of $\mathbf{P}_{AE}$ (measured by the terminating microphone in the artificial ear) to $\mathbf{P}_U$ in each of the real ears, for the same broadband drive signal level to the sound source. Each of these $\mathbf{P}_{AE}/\mathbf{P}_U$ curves can be thought of as an earphone calibration in that they indicate how an earphone drive voltage must be set to deliver a sound pressure stimulus to an experimental ear that is constant in level across frequency. Below 40 kHz, $\mathbf{P}_{AE}/\mathbf{P}_U > 0$ dB, which means that the artificial ear calibration would call for a lower stimulus than is needed to produce the desired sound pressure level in the real ear. Between 40 and 60 kHz, $\mathbf{P}_{AE}/\mathbf{P}_U < 0$ dB, which means that the artificial ear calibration would call for a larger stimulus than is needed to produce the desired sound pressure level in the real ear. Throughout the

frequency range there are ripples in magnitude and phase introduced by a mismatch in acoustic length between the artificial ear and the real ear. Together these factors introduce a substantial frequency-dependent bias in artificial ear calibrations across a wide frequency range.

Such bias will clearly influence high-frequency measurements that use $\mathbf{P}_U$ as an input, e.g., middle-ear transmission or auditory bandwidth. Quantities such as stapes velocity or cochlear or neural responses, when normalized by ear canal sound pressure, would appear artificially low by 5–20 dB below 40 kHz and artificially high by 5–15 dB in the 40- to 60-kHz range. For these reasons, artificial ear calibrations should be used with caution.

The similarity of $\Delta\mathbf{P}_{EC}$ at the ear canal entrance among the five ears (from two laboratories) in Fig. 7 suggests that a simple average EC sound pressure correction is possible. The variation in $\Delta\mathbf{P}_{EC}$ in Fig. 7 is less than $\mathbf{P}_{AE}/\mathbf{P}_U$ in Fig. 14, so an average ear canal correction leads to a more accurate estimation of middle-ear input than an artificial ear calibration.

VI. SUMMARY AND CONCLUSIONS

(1) The sound field in the gerbil bony ear canal can be described as (a) primarily uniform below 30 kHz and (b) a combination of uniform traveling and standing plane waves at higher frequencies. $|\mathbf{P}_{EC}|$ variations along the ear canal were on the order of 15 dB above 35 kHz and at least 20 dB above 60 kHz. Transverse variations across the ear canal at the TM were much smaller: about 2 dB.

(2) Sound pressure could be measured sufficiently near the umbo that no frequency-dependent notches in ear canal sound pressure were seen there over the frequency range of measurements (0.1–82 kHz). The result that variations across the TM were small means that the middle-ear input can be specified as sound pressure measured near the umbo at frequencies up to at least 60 kHz.

(3) The probe tube used to measure the sound pressure variations in some ear canals occluded ~30% of the ear canal but had a reasonably small effect on sound pressure near the umbo up to at least 60 kHz.

(4) Sound power reflectance computed from standing wave patterns indicates that 40%–80% of incident sound power at the TM enters the middle ear below 50 kHz, comparable to $\mathfrak{R}$ in cat, chinchilla, and human at lower frequencies. The ear canal power reflectance increases at higher frequencies. This increase in $\mathfrak{R}_{EC}$ as frequency increases is consistent with previous results in chinchilla and human and some previous results in cat, both at lower frequencies, and is consistent with the idea that the boundaries of the audiometric range are influenced by sound transmission into and through the middle ear.

(5) Sound pressure measured near the umbo in a real ear differed from the sound pressure measured at the termination of an artificial ear. If $\mathbf{P}_{AE}$ is used to set ear canal sound stimulus levels, these differences would bias stimulus levels downward below 40 kHz and upward between 40 and 60 kHz. These systematic biases throw into question the validity of the artificial ear calibration technique at high frequencies.

[1] For the simplest case of a uniform plane wave traveling along the ear canal, the deviation from maximum pressure is about 2 dB at $L=\lambda/10$, 3 dB at $\lambda/8$, and 6 dB at $\lambda/6$.

[2] Although it is not apparent because of the normalization of sound pressure in the figure, it is notable that if $B=A$ (corresponding to complete reflection), the pressure maximum will be twice as large as it will be if $B=0$ (corresponding to zero reflection).

[3] Transverse wave amplitude decays with distance x as $e^{-\kappa x}$. For the first antisymmetric mode, $\kappa d = 1.84$ (e.g., Fletcher, 1992, p. 198).

[4] Later experiments also used buprenex (0.2 mg/kg) at 6-h intervals as an analgesic.

[5] The sound field in the calibration coupler is a standing wave pattern with pressure reflection coefficient ~1—see Sec. V D.

[6] The reference location near the umbo in this ear may have been deeper into the EC than in the ears measured at MEEI.

[7] If the impedance of the TM is purely resistive or much larger than Z_0, $\theta = 0$; if the TM impedance is stiffness-dominated, $\theta > 0$ (the ear canal appears longer, e.g., Voss and Allen, 1994); and if the TM impedance is mass-dominated, $\theta < 0$ (the ear canal appears shorter).

[8] Acoustic impedance of the microphone $\mathbf{Z}_{mic}$ was computed from its equivalent volume $V_{mic} = 0.1$ mm^3 (G.R.A.S., 2006) by $\mathbf{Z}_{mic} = -j\ \rho c^2/(\omega V_{mic})$ (Beranek, 1986, p. 129).

[9] There are alternative methods for computing and checking $\mathfrak{R}$: (1) Stinson (1985a) showed that $\mathfrak{R}$ in a uniform tube could also be computed from the maximum slope of $\angle\Delta\mathbf{P}_{EC}$ with position $\Phi_x = \partial\angle\Delta\mathbf{P}_{EC}/\partial x$, which occurs where $\angle\Delta\mathbf{P}_{EC}$ crosses 0.25 or 0.75 cycles (at the same frequencies as magnitude notches). If the cross section of the tube decreases linearly toward the termination (a conical segment), the sound pressure magnitude increases near the termination relative to other locations in the tube and Φ_x gives a better estimate of $\mathfrak{R}$ than SWR. (2) The equivalence in standing wave patterns between frequency and location discussed in Sec. II means that $\mathfrak{R}$ could be computed from the slope of $\angle\Delta\mathbf{P}_{EC}$ with frequency $\Phi_f = \partial\angle\Delta\mathbf{P}_{EC}/\partial f$ equivalently to Φ_x, as long as $\mathfrak{R}$ varies slowly with frequency. (3) We performed a numerical parametric study of the relationship between SWR and the bandwidth BW of the spectral $\angle\Delta\mathbf{P}_{EC}$ transition from 0.125 to 0.375 cycles or from 0.625 to 0.875 cycles [defined as BW $= (f_{0.375} - f_{0.125})]/f_{0.25}$; similar for 0.625–0.875 cycle transition] using Eqs. (30) and (32) of Stinson (1985a). This study showed a simple inverse relationship between SWR and phase BW over the range $0.1 < A/B < 0.99$. We chose to compute $\mathfrak{R}$ from SWR rather than phase slope because (1) our measurements do not have the fine spatial resolution required for an accurate computation of Φ_x and (2) small local variations in Φ_f made the computation problematic. Figure 5(c) shows that, at frequencies as high as 65 kHz, $|\mathbf{P}_{EC}|$ near the bony ear canal entrance (distal to the first $|\mathbf{P}_{EC}|$ notch) was similar to $|\mathbf{P}_{EC}|$ near the umbo (between the first $|\mathbf{P}_{EC}|$ notch and the EC termination); therefore, errors in SWR due to an increase in $|\mathbf{P}_{EC}|$ caused by any reduction in EC cross-section area [as illustrated in Fig. 3 of Stinson, (1985a)] are likely to be small. (3) A comparison of $\angle\Delta\mathbf{P}_{EC}$ transition bandwidth (from Fig. 5) to $\angle\Delta\mathbf{P}_{AE}$ transition bandwidth (from Fig. 10), using a simple linear fit to $\angle\Delta\mathbf{P}_{EC}$ between 0.125 and 0.375 cycles, supports the idea that $\mathfrak{R}_{EC} < \mathfrak{R}_{AE}$ at middle frequencies and $\mathfrak{R}_{EC} \approx \mathfrak{R}_{AE}$ at high frequencies. (This method is less sensitive to the phase wiggles in Fig. 5.)

Allen, J. B. (**1986**). "Measurement of eardrum acoustic impedance," in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer, New York), pp. 44–51.

Ando, Y. (**1968**). "The directivity and the acoustic center of a probe tube microphone," J. Acoust. Soc. Jpn. **24**, 335–342.

Benade, A. H. (**1990**). *Fundamentals of Musical Acoustics* (Dover, New York).

Beranek, L. L. (**1986**). *Acoustics* (Acoustical Society of America, Melville, NY).

Chan, J. C. K., and Geisler, C. D. (**1990**). "Estimation of eardrum acoustic

pressure and of ear canal length from remote points in the canal," J. Acoust. Soc. Am. **87**, 1237–1247.

Cooper, N. P. (**2000**). "Radial variations in the vibrations of the cochlear partition," in *Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Takasaka, K. Ikeda, K. .. Ohyama, and T. Koike (World Scientific, Singapore), pp. 109–115.

Dallos, P. (**1996**). "Overview: Cochlear neurobiology," in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 1–43.

Decraemer, W. F., de la Rochefoucault, O., Dong, W., Khanna, S. M., Dirckx, J. J., and Olson, E. S. (**2006**). "Scala vestibuli pressure and three-dimensional stapes velocity measured in direct succession in gerbil," J. Acoust. Soc. Am. **121**, 2774–2791

Dong, W., and Olson, E. S. (**2006**). "Middle-ear forward and reverse transmission in gerbil," J. Neurophysiol. **95**, 2951–2961.

Fletcher, N. H. (**1992**). *Acoustic Systems in Biology* (Oxford University Press, New York).

Furlong, C., Ravicz, M. E., Rodgers, M. T., and Rosowski, J. J. (**2006**). "Real-time opto-electronic holographic measurements of the sound-induced displacement of tympanic membranes," *Abstracts of the 29th Midwinter Meeting of the Association for Research in Otolaryngology* (ARO, Mt. Royal, NJ), No. 644.

G. R. A. S. Sound and Vibration, Holte, Denmark (**2006**), http://www.grasinfo.dk/documents/pd_40DP_ver_04_09_06.pdf, viewed 29 December 2006.

Hu, A., Cuomo, F. W., and Zuckerwar, A. J. (**1992**). "Theoretical and experimental study of a fiber optic microphone," J. Acoust. Soc. Am. **91**, 3049–3056.

Huang, G. T., Rosowski, J. J., Puria, S., and Peake, W. T. (**2000**). "A noninvasive method for estimating acoustic admittance at the tympanic membrane," J. Acoust. Soc. Am. **108**, 1128–1146.

Keefe, D. H., and Benade, A. H. (**1981**). "Impedance measurement source and microphone proximity effects," J. Acoust. Soc. Am. **69**, 1489–1495.

Keefe, D. H., Bulen, J. C., Arehart, K. H., and Burns, E. M. (**1993**). "Ear-canal impedance and reflection coefficient in human infants and adults," J. Acoust. Soc. Am. **94**, 2617–2638.

Khanna, S. M., and Stinson, M. R. (**1985**). "Specification of the acoustical input to the ear at high frequencies," J. Acoust. Soc. Am. **77**, 577–589.

Khanna, S. M., and Tonndorf, J. (**1972**). "Tympanic membrane vibrations in cats studied by time-averaged holography," J. Acoust. Soc. Am. **51**, 1904–1920.

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**1982**). *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York).

Lawton, B. W., and Stinson, M. R. (**1986**). "Standing wave patterns in the human ear canal used for estimation of acoustic energy reflectance at the eardrum," J. Acoust. Soc. Am. **79**, 1003–1009.

Lay, D. M. (**1972**). "The anatomy, physiology, functional significance and evolution of specialized hearing organs of gerbilline rodents," J. Morphol. **138**, 41–120.

Lynch, T. J., III (**1981**). "Signal processing by the cat middle ear: Admittance and transmission, measurements and models," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge.

Lynch, T. J. III, Peake, W. T., and Rosowski, J. J. (**1994**). "Measurements of the acoustic input impedance of cat ears, 10 Hz to 20 kHz," J. Acoust. Soc. Am. **96**, 2184–2209.

Magnusson, P. C. (**1965**). *Transmission Lines and Wave Propagation* (Allyn and Bacon, Boston).

Maki, K., and Furukawa, S. (**2005**). "Acoustical cues for sound localization by the Mongolian gerbil, Meriones unguiculatus," J. Acoust. Soc. Am. **118**, 872–886.

Margolis, R. H., Paul, S., Saly, G. L., Schachern, P. A., and Keefe, D. H. (**2001**). "Wideband reflectance tympanometry in chinchillas and humans," J. Acoust. Soc. Am. **110**, 1453–1464.

Margolis, R. H., Saly, G. L., and Keefe, D. H. (**1999**). "Wideband reflectance tympanometry in normal adults," J. Acoust. Soc. Am. **106**, 265–280.

Morse, P., and Ingard, U. (**1968**). *Theoretical Acoustics* (McGraw-Hill, New York), pp. 509–514.

Neely, S. T., and Gorga, M. P. (**1998**). "Comparison between intensity and pressure as measures of sound level in the ear canal," J. Acoust. Soc. Am. **104**, 2925–2934.

Olson, E. S. (**1998**). "Observing middle and inner ear mechanics with novel intracochlear pressure sensors," J. Acoust. Soc. Am. **103**, 3445–3463.

Olson, E. S. (**2001**). "Intracochlear pressure measurements related to cochlear tuning," J. Acoust. Soc. Am. **110**, 349–367.

Olson, E. S., and Cooper, N. P. (**2000**). "Stapes motion and scala vestibuli pressure in gerbil," *Abstracts of the 23rd Midwinter Meeting of the Association for Research in Otolaryngology* (ARO, Mt. Royal, NJ), No. 399.

Overstreet, E. H., Richter, C.-P., Temchin, A. N., Cheatham, M. A., and Ruggero, M. A. (**2003**). "High-frequency sensitivity of the mature gerbil cochlea and its development," Audiol. Neuro-Otol. **8**, 19–27.

Overstreet, E. H., and Ruggero, M. A. (**2002**). "Development of wide-band middle ear transmission in the Mongolian gerbil," J. Acoust. Soc. Am. **111**, 261–270.

Pearce, M., Richter, C.-P., and Cheatham, M. A. (**2001**). "A reconsideration of sound calibration in the mouse," J. Neurosci. Methods **106**, 57–67.

Pierce, A. D. (**1981**). *Acoustics: An Introduction to its Physical Principles and Applications* (Acoustical Society of America, Melville, NY).

Puria, S., and Allen, J. B. (**1998**). "Measurements and model of the cat middle ear, Evidence of tympanic membrane delay," J. Acoust. Soc. Am. **104**, 3463–3481.

Rabbitt, R. D., and Holmes, M. H. (**1988**). "Three-dimensional acoustic waves in the ear canal and their interaction with the tympanic membrane," J. Acoust. Soc. Am. **83**, 1064–1080.

Ravicz, M. E. (**1990**). "Acoustic impedance of the gerbil ear," M.S. thesis, Boston University, Boston, MA.

Ravicz, M. E., and Rosowski, J. J. (**2004**). "High-frequency sound transmission through the gerbil middle ear," *Abstracts of the 27th Midwinter Meeting of the Association for Research in Otolaryngology* (ARO, Mt. Royal, NJ), No. 817.

Ravicz, M. E., Rosowski, J. J., and Voigt, H. F. (**1992**). "Sound-power collection by the auditory periphery of the mongolian gerbil Meriones unguiculatus. I. Middle-ear input impedance," J. Acoust. Soc. Am. **92**, 157–177.

Ravicz, M. E., Rosowski, J. J., and Voigt, H. F. (**1996**). "Sound-power collection by the auditory periphery of the mongolian gerbil Meriones unguiculatus. II. External-ear radiation impedance and power collection," J. Acoust. Soc. Am. **99**, 3044–3063.

Ren, T., and Nuttall, A. L. (**2001**). "Basilar membrane vibration in the basal turn of the sensitive gerbil cochlea," Hear. Res. **151**, 48–60.

Rosowski, J. J. (**1992**). "Hearing in transitional mammals: Predictions from the middle-ear anatomy and hearing capabilities of extant mammals," in *The Evolutionary Biology of Hearing*, edited by D. B. Webster, R. R. Fay, and A. N. Popper (Springer, New York), pp. 615–631.

Rosowski, J. J., Carney, L. H., Lynch, T. J., III, and Peake, W. T. (**1986**). "The effectiveness of external and middle ears in coupling acoustic power into the cochlea," in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer, New York), pp. 3–12.

Rosowski, J. J., Carney, L. H., and Peake, W. T. (**1988**). "The radiation impedance of the external ear of cat: Measurements and applications," J. Acoust. Soc. Am. **84**, 1695–1708.

Rosowski, J. J., Davis, P. J., Merchant, S. N., Donahue, K. M., and Coltrera, M. D. (**1990**). "Cadaver middle ears as models for living ears: Comparisons of middle ear input immittance," Ann. Otol. Rhinol. Laryngol. **99**, 403–12.

Rosowski, J. J., Furlong, C., Ravicz, M. E., and Rodgers, M. T., (**2006**). "Real-time opto-electronic holographic measurements of sound-induced tympanic membrane displacements," 4th International Symposium on Middle Ear Mechanics in Research and Otolaryngology, Zürich, Switzerland, No. P14.

Rosowski, J. J., Ravicz, M. E., Teoh, S. W., and Flandermeyer, D. (**1999**). "Measurements of middle-ear function in the Mongolian gerbil, a specialized mammalian ear," Audiol. Neuro-Otol. **4**, 129–136.

Ruggero, M. A., and Temchin, A. N. (**2002**). "The roles of the external, middle, and inner ears in determining the bandwidth of hearing," Proc. Natl. Acad. Sci. U.S.A. **99**, 13206–13210.

Shaw, E. A. G. (**1974**). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," J. Acoust. Soc. Am. **56**, 1848–1860.

Siegel, J. H. (**1994**). "Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes," J. Acoust. Soc. Am. **95**, 2589–2597.

Sivian, L. J., and White, S. D. (**1933**). "On minimum audible fields," J. Acoust. Soc. Am. **4**, 288–321.

Stevens, K. N., Berkovitz, R., Kidd, G., and Green, D. M. (**1987**). "Calibration of ear canals for audiometry at high frequencies," J. Acoust. Soc. Am. **81**, 470–484.

Stinson, M. R. (**1985a**). "Spatial variation of phase in ducts and the mea-

surement of acoustic energy reflection coefficients," J. Acoust. Soc. Am. **77**, 386–393.

Stinson, M. R. (**1985b**). "The spatial distribution of sound pressure within scaled replicas of the human ear canal," J. Acoust. Soc. Am. **78**, 1596–1602.

Stinson, M. R. (**1990**). "Revision of estimates of acoustic energy reflectance at the human eardrum," J. Acoust. Soc. Am. **88**, 1773–1778.

Stinson, M. R., and Khanna, S. M. (**1989**). "Sound propagation in the ear canal and coupling to the eardrum, with measurements on model systems," J. Acoust. Soc. Am. **85**, 2481–2491.

Stinson, M. R., and Khanna, S. M. (**1994**). "Spatial distribution of sound pressure and energy flow in the ear canals of cats," J. Acoust. Soc. Am. **96**, 170–180.

Stinson, M. R., Shaw, E. A. G., and Lawton, B. W. (**1982**). "Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal," J. Acoust. Soc. Am. **72**, 766–773.

Voss, S. E., and Allen, J. B. (**1994**). "Measurement of acoustic impedance and reflectance in the human ear canal," J. Acoust. Soc. Am. **95**, 372–384.

Wever, E. G., and Lawrence, M. (**1954**). *Physiological Acoustics* (Princeton Univiversity Press, Princeton, NJ).

Zwislocki, J., Kruger, B., Miller, J. D., Niemoller, A. F., Shaw, E. A. G., and Studebaker, G. (**1988**). "Earphones in audiometry," J. Acoust. Soc. Am. **83**, 1688–1689.

Effects of mobile phone exposure on time frequency fine structure of transiently evoked otoacoustic emissions

Alessia Paglialonga[a]
Institute of Biomedical Engineering ISIB, Italian National Research Council, Milan, Italy; and Department of Biomedical Engineering, Polytechnic University of Milan, Milan, Italy

Gabriella Tognola and Marta Parazzini
Institute of Biomedical Engineering ISIB, Italian National Research Council, Milan, Italy

Mark E. Lutman and Steven L. Bell
Institute of Sound and Vibration Research, University of Southampton, Southampton, England

Gyorgy Thuroczy
National Research Institute for Radiobiology and Radiohygiene NIRR, Department of Non-Ionizing Radiation, Budapest, Hungary

Paolo Ravazzani
Institute of Biomedical Engineering ISIB, Italian National Research Council, Milan, Italy

Mobile phones have become very commonly used worldwide within a short period of time. To date there is only limited knowledge about interaction between electromagnetic fields (EMFs) emitted by mobile phones and the auditory function. Moreover, there is widespread concern that there may be potential for harm. The aim of this study was to assess potential subtle changes in cochlear function by measuring the temporal and spectral fine structure of transiently evoked otoacoustic emissions (TEOAE) in normal hearing subjects after exposure to EMFs emitted by Global System for Mobile Communication (GSM) mobile phones. TEOAEs were recorded in 27 healthy young adults before and after 10 min of real or sham exposure in a double-blind design. TEOAE data were analyzed both globally (broadband analysis) and using the Wavelet Transform (analysis of the time-frequency fine structure). The broadband analysis revealed no significant effect on TEOAEs related to exposure, confirming results of previous studies; in addition, no significant change was detected in the analysis of the elementary wavelet components, suggesting that the temporal and spectral fine structure of TEOAEs is not affected by 10 min exposure to low-intensity EMFs emitted by GSM mobile phones. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2773944]

I. INTRODUCTION

The rapid worldwide increase in the use of mobile phones raises questions about the possible adverse effects of emitted electromagnetic fields (EMFs) on health. To date, biological effects of radio-frequency (rf) EMFs transmitted by these devices are still a matter of public and scientific discussion. Sensations of burning or warmth around the ear (Oftedal *et al.*, 2000), headache (Frey, 1998), disturbance of sleep (Borbely *et al.*, 1999), alteration of cognitive functions and neural activity (Preece *et al.*, 1999), as well as alteration of the blood-brain barrier (Fritze *et al.*, 1997) and a relative decrease in regional cerebral blood flow (Haarala *et al.*, 2003), have been reported as effects resulting from mobile telephone use. The health implications of mobile phones and base stations were reviewed in a report issued by the English National Radiological Protection Board (NRPB) advisory group on nonionizing radiation (Documents of the NRPB, 2003). This report concluded that "*the weight of evidence now available does not suggest that there are adverse health effects from exposures to rf fields below guideline levels, but the published research on rf exposures and health has limitations, and mobile phones have only been in widespread use for a relatively short time. The possibility therefore remains open that there could be health effects from exposure to rf fields below guideline levels; hence continued research is needed.*"

Of all major anatomical structures, the ear is in closest proximity to the mobile phone, resulting *a priori* in a fairly high absorption rate of the incident EMF (Burkhardt and Kuster, 2000). This may lead to higher energy deposition in the cochlea compared to other parts of the body. However, there is still only limited knowledge on potential adverse effects of EMF on hearing, particularly when a low level of power, such as that used in GSM (Global System for Mobile Communication) mobile phone technology, is considered. In fact, little has been published on effects of EMFs on the ear as a sense organ and results are not always consistent. This is notable when considering the proximity of the ear to the radiation source and the sensitivity of its sensory cells to

---

[a]Author to whom correspondence should be addressed. Electronic mail: alessia.paglialonga@polimi.it

exogenous and endogenous agents, and it is easy to imagine a high impact on the peripheral auditory system.

Dealing with central auditory pathways, some authors studied the interactions between mobile phones and the central hearing system in human subjects. In particular, Kellenyi et al. (1999) found that 15 min exposure to GSM phone radiation caused an increase in auditory brainstem response (ABR) when measured in the exposed side of the head; on the contrary, Arai et al. (2003) demonstrated that 30 min mobile phone use had no short-term adverse effects on ABR recordings, ABR recovery function and middle latency response; Hamblin et al. (2004) measured event-related potentials, showing that exposure might have affected neural activity, particularly in proximity to the phone, but recognized that caution should be applied in interpreting results due to the small sample size used in that experiment.

Dealing with the inner ear, the cochlear outer hair cells (OHCs) functionality can be assessed by means of otoacoustic emissions (OAE), a noninvasive physiological measure which is known to directly reflect OHC function (Probst et al., 1991). OHCs are special sensory cells that enhance cochlear sensitivity and frequency selectivity by amplifying low-level sound signals mechanically (Dallos, 1992). In vitro, OHCs are capable of fast contractions and elongations of their cell body in response to an electric field (Brownell et al., 1985). This electromotility is assumed to produce the amplification of vibrations in the cochlea during acoustic stimulation (Zheng et al., 2000), and can be directly assessed by means of OAEs. The high test-retest reliability of OAE measures allows the utilization of these emissions, and in particular distortion product otoacoustic emissions (DPOAE) and transiently evoked otoacoustic emissions (TEOAE), for monitoring dynamic changes in cochlear responsiveness (Lonsbury-Martin and Martin, 1990); this is even more true when, as shown by Janssen et al. (2005) for DPOAEs, high signal-to-noise ratio (SNR) is achieved. Moreover, OAEs are a well-described method for detecting cochlear involvement by external agents. Low-level noise exposure (Skellett et al., 1996), increased body temperature due to fever (O'Brien, 1994), administration of salicylate (McFadden and Plattsmier, 1984; Janssen et al., 2000; Parazzini et al., 2005a), or alteration of body posture (Fukai et al., 2005) are known to considerably affect OAE amplitude.

Some authors studied the interaction between mobile phone radiation and the cochlear OHCs functionality by means of OAEs both in animals and humans, observing sometimes significant effects that might be attributable to exposure, but these results are inconsistent and need further confirmation. Marino et al. (2000) found no statistically significant evidence measuring DPOAEs in Sprague-Dawley rats after chronic exposure to EMFs. Kizilay et al. (2003) measured DPOAEs in adult and developing rats after chronic exposure to EMFs from a mobile phone showing no hearing deterioration, at least at outer ear, middle ear, and cochlear levels. Grisanti et al. (1998) found that modulated EMFs from Total Access Communication System (TACS) phones modified the amplitude (to the amount of about 1 dB) of DPOAEs in humans. Ozturan et al. (2002), measuring DPOAEs and TEOAEs after 10 min exposure to EMFs from

a mobile phone, showed that EMFs had no immediate effect at cochlear level. Janssen et al. (2005) observed, in some of the human subjects involved in their study, an extremely small EMF-exposure-correlated change in DPOAE level (<1 dB), which could be considered, as suggested by the authors, physiologically irrelevant with respect to a potential health risk. Taken together, these results appear contradictory: only sporadic effects were found and results from the different studies overlap only to a very limited extent despite the fact that similar experimental conditions were used.

As a consequence of the limited evidence concerning mobile phone exposure and its influence on the hearing system, in 2002 the European Commission established the Project GUARD: "Potential Adverse Effects of GSM Cellular Phones on Hearing" that lasted until 2004. This research program focused on the potential effects of GSM cellular phones on the hearing system of laboratory animals and humans. In the course of the GUARD project, additional to the main audiological measures (i.e., pure tone audiometry, tympanometry, and ABR) DPOAEs and TEOAEs were measured in order to assess possible acute changes in OHC function as a consequence of exposure to low-intensity EMFs transmitted by GSM cellular phones at frequencies of 900 and 1800 MHz (see, e.g., Bell et al., 2004; Ravazzani et al., 2005). Some results obtained in the course of the Project GUARD, on both laboratory animals and humans, were recently published. Results from animal studies (Galloni et al., 2005a; Galloni et al., 2005b; Parazzini et al., 2007a) clearly showed the lack of effects of EMFs after subchronic exposures on DPOAEs of rats, both in normal ears as well as in ears exposed to a well recognized pathological agent (i.e., gentamicin). Preliminary results from human studies reported no significant effect on DPOAEs (Parazzini et al., 2005b) as well as on hearing threshold levels (HTLs) and TEOAEs (Uloziene et al., 2005) due to exposure to microwaves at 900 and 1800 MHz, as confirmed also by the pooled analysis of the data obtained in the course of the GUARD project, which showed no effects of GSM mobile phones on HTL, ABR, TEOAEs, and DPOAEs (Parazzini et al., 2007b).

In all the abovementioned studies assessing possible influence of EMF exposure on the inner ear, the amplitude of DPOAEs and TEOAEs was considered as the only relevant parameter for the quantitative comparison of OAEs measured before and after exposure to GSM EMFs, and TEOAE data were studied either in the whole emission frequency band or in a number of 1-kHz-wide bands.

The aim of the present study was to assess possible effects of EMFs on transiently evoked otoacoustic emissions (TEOAEs) measured on a group of normal-hearing subjects by using a high resolution time-frequency approach and analyzing OAE energy and an additional relevant parameter related to the temporal structure of the signal (i.e., latency) in order to examine whether and to what extent the temporal and spectral features of TEOAE fine structure could be affected by the emitted radiation.

TEOAEs, which are responses to clicks, are nonstationary signals comprising cumulative responses from the entire cochlea. In order to separate the contributions from the dif-

ferent frequencies, several methods for time-frequency analysis of TEOAEs have been proposed (Cheng, 1995; Jedrzejczak *et al.*, 2004). Among these methods for the analysis of time-varying signals in the time-frequency domain, the wavelet transform (WT) is frequently used because its time and frequency resolutions are suitable for accurately determining both the frequency content and the time course (e.g., the latency) of TEOAEs (Marozas *et al.*, 2006; Moleti *et al.*, 2005; Yang *et al.*, 2003; Tognola *et al.*, 1998). This frequency-specific analysis of TEOAE recordings, separating the different frequency contributions and preserving a high temporal resolution, might improve TEOAE precision in detecting even small changes in outer hair cell function due to EMF exposure.

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Human Experimentation Safety and Ethics Committee of the Institute of Sound and Vibration Research, which was in charge for human experimentation phases.

II. METHODS

A. Subjects

A population of 29 otologically normal subjects took part in the experiment. Participants were healthy young adults (23–30 years old; 17 male, 12 female) without any evidence of hearing or ear disorders, thus maximizing the sensitivity of the study to detect small changes that might occur. Acceptance as participants was based on otoscopy, audiometry by air conduction (0.5, 1, 2, 3, 4, 6, and 8 kHz) and bone conduction (0.5, 1, and 2 kHz), tympanometry, acoustic reflex testing [1 kHz tone at 100 dB hearing level (HL)], and a simple screening questionnaire concerning medical and otological history. Audiometric hearing thresholds of 20 dB HL or better at each of the test frequencies were required.

B. Experimental protocol and TEOAE measurement

The experimental paradigm used in this study was a within-subject design. This entailed TEOAE measurements immediately before and immediately after 10 min exposure to EMFs via a commercially available mobile phone, with the OAE probe removed during the exposure. No controlled pauses were included between the registration and the exposure phases. The procedure was conducted twice in a double-blind design: once with a *real* (test) exposure and once with a *sham* (control) exposure. This approach maximizes sensitivity to change because between-subject variation is minimized by calculating the difference between before and after measurements. Real and sham exposures were on separate days (at least 24 h apart) with both the test participant and tester being blind to the condition being used. Practical and ethical considerations impose severe restraints on the intensity and duration of exposure that can be allowed. Therefore, exposure dose was necessarily low (see also below for additional details on the exposure setup), but comparable with the use of a mobile phone in normal day life.

TEOAE measurements were performed only on the exposed ear. The selection of the test ear was based on the audiometric and audiological data obtained from the screening session. The principle was to select the ear with the better results on the main outcome measures (TEOAE, DPOAE, and ABR). This approach was designed to give maximum sensitivity to detect any change.

The Otodynamics ILO-88 system was used to record TEOAEs according to the so-called "linear" protocol, using two trains of click stimuli (80 μs rectangular clicks presented at a rate of 50/s) at 60, 70, and 80 dB peak sound pressure level (SPL). The rationale of using the lower click intensities (i.e., 60 and 70 dB) was to maximize sensitivity to change because OAEs are reported to be most sensitive to cochlear impairment when low eliciting stimuli are used (Kummer *et al.*, 1998). Responses to the click trains were stored and averaged in two 20 ms buffers A and B. Each measurement run included a minimum of 500 sweeps (i.e., 2000 clicks). For investigating minute changes in auditory function, high measurement accuracy is needed. Only subjects exhibiting high emission levels above the noise floor were included in this study: in particular, a TEOAE was defined as a real evoked response if the SNR was more than 3 dB and, in addition, SNR greater than 6 dB in two or more half-octave bands centered at 1.5, 2, 3, and 4 kHz was required. This choice is also in agreement with the consideration that, since SNR directly influences measurement accuracy (Janssen *et al.*, 2005), detection of possible minute changes in TEOAEs induced by EMF exposure requires high SNRs. Twenty-seven out of the 29 subjects completed the experimental procedure and had clear recordable TEOAEs satisfying the SNR criteria stated above.

All testing was carried out in a sound-treated room satisfying criteria in ISO 8253-1 for air conduction audiometry.

C. Exposure setup and positioning system

For the EMF exposure, the normal output of a GSM mobile phone (NOKIA 6310), connected by serial cable to a PC, was used at full power for 10 min. Half of the participants received GSM exposure at 900 MHz (full emitted power=2 W) and the other half received GSM exposure at 1800 MHz (full emitted power=1 W). The mobile phone was controlled via special software provided by NOKIA to set the exposure parameters to the required frequency (900 or 1800 MHz) and required power.

The sham or real exposure was realized using a "load" or a "dummy load." For this purpose an external power load was connected to the remote antenna connector of the phone. A 50 Ω resistive load and an open-circuit dummy load were developed for sham/exposed conditions with the same shape and structure. In order to confirm the effectiveness of the load, surface scanning of the phone by near field measurement was performed. The load intercepts the rf signal to the internal antenna on the phone and dissipates the rf in the load, while the dummy load looks identical but does nothing, allowing the rf to reach the antenna. No radiated rf fields were measured using the load connected to the external antenna output, confirming its effectiveness for the sham exposure. Local over-heating of the load with respect to the dummy load could not be detected by the tested subjects

Paglialonga *et al.*: Mobile phone effects on otoacoustic emissions

because the load was connected to the back of the phone, with no direct contact with any part of subject body.

In order to estimate the level of EMF exposure in the inner ear region, measurements of absorbed radio frequency power in the head, termed Specific Absorption Rate [SAR (W/kg)], were made in standard brain tissue equivalent liquid in a phantom (Antennessa GEL-900/1800, France). In these measurements, a three-dimensional step motor robot system (Arrick Robotics, 3-Axis Positioning Table, USA), internal E-field probe (ER3DV4R, Schmid & Partner Engineering Ag., Switzerland) and a nonmetallic phone positioning system were applied; the phone was held on the phantom in the "touch position," according to CENELEC standard EN 50361. These measurements, at a distance of 30 mm from the surface of the phantom which corresponds approximately to that of the cochlea, revealed maximum SAR of 0.41 and 0.19 W/kg for the 900 MHz and 1800 MHz frequencies, respectively. These values are below 2 W/Kg which is the limit of the European Council recommendation 1999/519/EC. The output of the phone had previously been measured using a frequency and power meter (Hewlett-Packard 5374A, USA) and circulators (RYT Industries, Santa Clara, USA) through the external antenna output of the phone. To satisfy our experimental protocol, the phones had to emit the same rf power throughout the course of the 10 min exposure. Since the phones have battery supply, the measurement of long term rf output stability was made by PC data acquisition of the output power during the whole discharge period of the battery. The phone rf power uncertainty during the whole period was below 1% using the highest power level at 900 MHz, (2 W) and at 1800 MHz (1 W). In the first 10 min the uncertainty was below 0.4% at both GSM frequencies.

In case there might be an interaction between acoustic and EMF exposure at the OHC level, a controlled sound exposure was introduced in both real and sham conditions. This consisted of running speech to which the participant was asked to attend.

To prevent any possible effects from using the speaker in the handset, the speech material was delivered via an ER-3A insert earphone. The insert earphone delivered sound via an extended polyethylene tube, which was taped along the participant's jaw with the end of the tube placed at the tragal notch. The speech signal had been pre-filtered to compensate for the altered frequency response of the earphone used in this way. Speech was presented at a typical conversational level, with the level of the sound at the eardrum equivalent to that produced by free field speech material at 60 dB(A).

A system of phone fixation that allowed free head movement by the user had been designed and developed for the GUARD project. The positioning holder has three main parts: a headband, an adjustable arm and a phone holder. All parts of the positioning system were made of nonmetallic plastic materials in order to avoid any perturbation of the EMF emitted by the mobile phone. The headband allowed free movement of the head without any disturbance of the phone from the adjusted position. By using the adjustable arm the phone may be placed as required and may be ad-

justed according to the size of the subject's head under investigation. The adjustable arm can be placed on either side of the headband. The phone holder was attached to a bracket glued to the battery pack of the phone, which is on the reverse side from the keypad. During the exposure the phone was placed so that its longitudinal axis followed an imaginary line from the entrance to the ear canal to the corner of the mouth, in accordance with the CENELEC standard EN 50361.

D. TEOAE data processing and parameter extraction

TEOAEs measured in presence or absence of EMF exposure, considering both the broadband signal (i.e., the whole frequency content of the emission) and its high resolution time-frequency representation, were compared on the basis of the measured energy, computed in the whole time window

$$EN = \sum_{i=1}^{N} \left(\frac{A_i + B_i}{2} \right)^2, \tag{1}$$

where A_i and B_i are the samples, at the ith time instant, of the TEOAE responses stored in the recording buffers A and B and N is the length of the time window.

The time-frequency representations of TEOAE data were obtained by means of the WT (Mallat, 1989), a technique for time-frequency analysis of time-varying signals (i.e., signals whose frequency content varies along with time). It is well established that TEOAEs show a "frequency dispersion" reminiscent of the place-frequency distribution along the cochlea (Kemp, 1979). The time of appearance of a component seen in a TEOAE is intimately related to the frequency of the component itself: the higher the frequency the shorter the latency, and vice versa, the lower the frequency the longer the latency. In the case of TEOAEs, it has been proved (Marozas *et al.*, 2006; Moleti *et al.*, 2005; Yang *et al.*, 2003; Tognola *et al.*, 1998) that the WT has both time and frequency resolutions suitable for accurately determining both the frequency content and the latency of a TEOAE component.

The WT at the generic time t and frequency f of signal $x(t)$ is defined as (Mallat, 1989)

$$WT(\tau, f) = \int_t x(t) \cdot \sqrt{f/f_0} \cdot \psi^*(f/f_0 \cdot (t - \tau)) dt, \tag{2}$$

where * denotes complex conjugation. Basically, the WT decomposes a signal $x(t)$ into elementary components by means of a bank of bandpass filters $\sqrt{f/f_0} \cdot \psi(f/f_0(t))$ that are iteratively derived from a unique prototype function $w(t)$, called the mother wavelet. The mother wavelet $w(t)$ is a function with finite energy centered around time $t=0$; its Fourier transform is a bandpass function centered around frequency f_0. The following mother wavelet, from the family of functions proposed by Wit *et al.* (1994) and by Tognola *et al.* (1997) was considered:

$$w_n(t) = \frac{1}{(1 + t^{2n})} \cos(20t), \quad n = 4. \tag{3}$$

The bandwidth of the bandpass filters derived from the mother wavelet is proportional to the central frequency of the filter, whereas the duration of their impulse response in the time domain is inversely proportional to the central frequency of the filter. As such, the WT is a "constant-Q filter bank" (where Q, the quality factor, is defined as the ratio between the center frequency and the bandwidth of a filter). As the WT is a spectral analyzer, it is suitable for analyzing both linear and nonlinear systems, such as the nonlinear components of TEOAE responses. In particular, the performance of the WT in processing TEOAEs seems to be intrinsically related to the fact that the cochlea itself behaves as a parallel bank of constant Q-factor bandpass filters (Yang *et al.*, 1992; Tognola *et al.*, 1998).

The set of values of $WT(\tau, f)$ calculated from Eq. (2) for each time τ and frequency f is the "time-frequency spectral distribution" of the signal $x(t)$ and gives an indication as to which spectral components are present and at which temporal intervals. By using Eq. (2), TEOAEs were decomposed into 98-Hz-wide frequency bands in the 0.83–4.06 kHz range for a total of 33 "elementary components" for each TEOAE recording. Frequencies below 0.83 kHz were not included in the analysis because of the low SNR, and frequencies higher than 4.06 kHz were excluded because in this range no relevant emissions can be measured in adult subjects (Morand *et al.*, 2000).

In addition to energy (EN), latency (LAT) of the 33 elementary TEOAE components was used to quantitatively evaluate the modifications of all the frequency-specific TEOAE contributions induced by EMF exposure. The latency of the elementary frequency component was defined, according to the definition proposed by Tognola *et al.* (1998), as the time interval from the stimulus onset to the maximum of the temporal envelope of the wave form.

In order to avoid analyzing highly noisy band-limited components, percent reproducibility (i.e., the zero-lag cross-correlation coefficient between A and B replicates) was measured and only those components where reproducibility was ≥50% were selected and included in the analysis, whereas the others were discarded. As a consequence, the number of subjects involved in the analysis varied (from 25 to 27 individuals) as a function of the frequency band.

E. Statistical analysis

The Kolmogorov-Smirnov test for normality and the D'Agostino test for skewness, performed on TEOAE parameters (i.e., EN and LAT), indicated that both distributions were neither normal nor symmetric. Thus, data were transformed to near normality in order to satisfy the assumptions required to perform analysis of variance. In order to maximize symmetry of the two parameters, EN was mapped with a logarithmic transformation and LAT with a square root transformation. The D'Agostino test for skewness revealed that in all the 33 frequency bands the distributions of EN and LAT were symmetric (α set at 5%) after logarithmic and square root transformation, respectively, so parametric analysis of variance was performed on the transformed data.

To test if there were any significant changes in TEOAE parameters due to GSM exposure, multivariate repeated measures analyses of variance were performed (α set at 5%). The experimental paradigm used in this study was a within-subject design. Thus, analyses of variance focused on within-subject variations rather than on differences between subjects. Exposure (real vs sham) and time of testing (before vs after exposure) were the two within-subject factors. The broadband data and each of the 33 elementary components were analyzed independently of each other.

III. RESULTS

During each recording, percent stimulus stability (i.e., an estimate of average correlation between the wave forms of the stimulus at the beginning and at the end of each recording) was checked by the tester. In the examined data set 93% of the recordings had stimulus stability above 90%, with a mean value of 95% at each of the three stimulus levels. In addition, changes in stimulus level in the paired measurements before and after exposure were negligible: the stimulus level, monitored by the intensity of the largest peak of meatal response, varied less than 2 dB between the two consecutive recordings in 95% of tests.

All subjects included in this analysis had very high TEOAE reproducibility, with an average percent value across the group of 98.73% and a standard deviation of 1.88%. Actually, reproducibility was almost always around 99%, with only two subjects having 85.5% and 89.8% at 60 dB SPL.

As to SNR, though the lower limit was set at 3 dB, much higher values were measured, up to 35 dB. Average SNR was 26, 21, and 16 dB at 80, 70, and 60 dB SPL, respectively, with standard deviations around 3 dB at each of the three stimulus levels.

An example of TEOAE data from one of the subjects that participated in the study is shown in Fig. 1. The measured emission wave form (broadband TEOAE) is shown, together with nine elementary components taken from the whole set of 33 elementary band-limited components (centered around 879, 977, 1075, 1173, 1271, 1369, 1467, 1565, and 1663 kHz) obtained after WT processing of TEOAE signals.

An example of parameters extracted from the band-limited TEOAE components, from one of the subjects that participated in this study, is shown in Fig. 2. Energy (EN) and latency (LAT) of the 33 elementary frequency components (spanning frequencies from 0.83 to 4.06 kHz) of TEOAE measured before and after exposure to EMF are shown.

A. Broadband TEOAEs

Figure 3 shows the mean and standard deviation of the shifts (i.e., the differences between the recordings before and after a real or sham exposure to a GSM mobile phone) for TEOAE energy (EN). The maximum mean shifts in EN were equal to -53 mPa^2 after a sham exposure and -25 mPa^2 after a real exposure for recording stimulus equal to 80 dB. Analy-
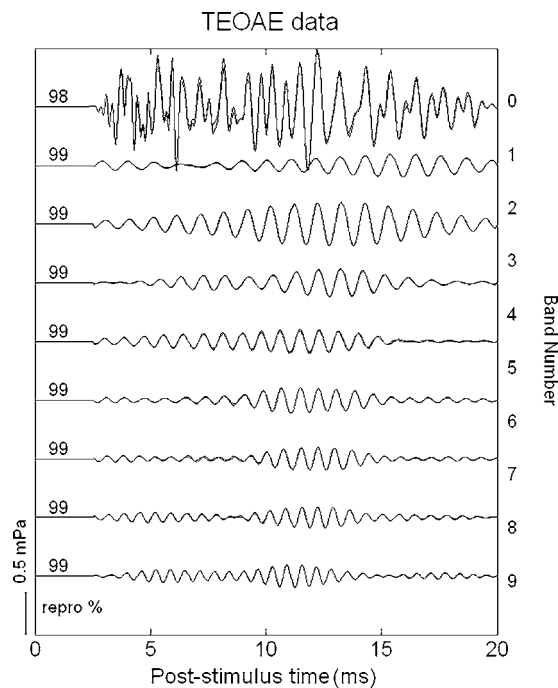
FIG. 1. Example of TEOAE data from one of the subjects that participated in the study. In the first row, broadband TEOAE (raw data) are shown, labeled as band '0,' whereas the other rows show the elementary band-limited components obtained with the WT. For clarity purposes only few components, i.e., the first 9 out of the 33, are shown, centered around 879, 977, 1075, 1173, 1271, 1369, 1467, 1565, and 1663 kHz. In each row, two replicate recordings (A and B in the ILO equipment) are superimposed. Numbers on the left are the reproducibility (percent value) between the two replicate recordings.
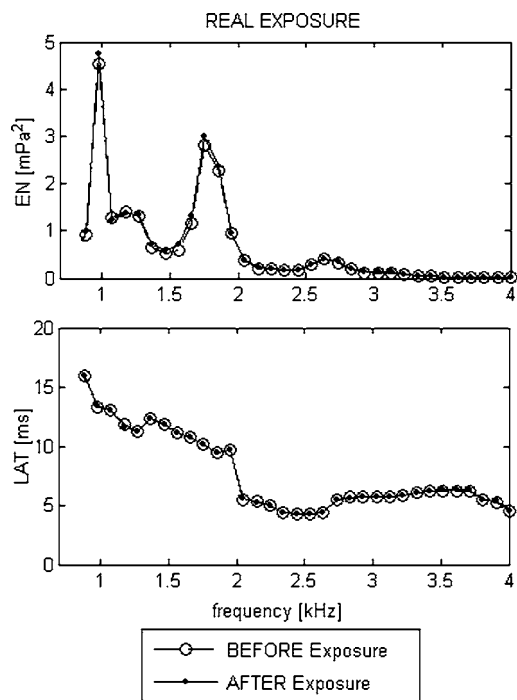
FIG. 2. Example of energy (EN, top panel) and latency (LAT, bottom panel), from one of the subjects that participated in the study, in the 33 elementary band limited components, spanning frequencies from 0.83 to 4.06 kHz, obtained after WT processing of TEOAE signals measured before and after exposure to EMFs.
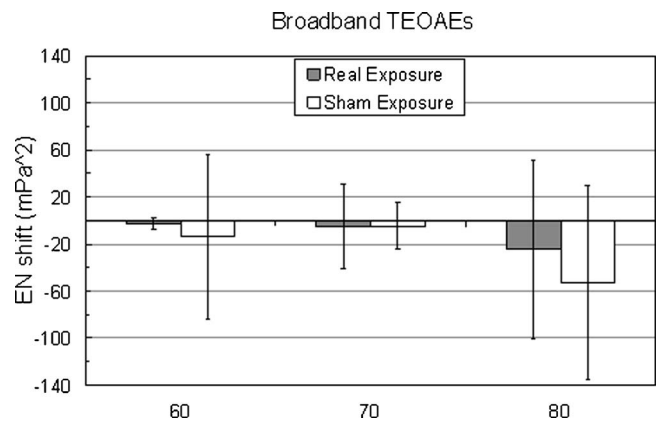
FIG. 3. Energy (EN) shifts in broadband TEOAEs after real or sham exposure to a GSM mobile phone with respect to measurement before exposure, for all the three stimulus levels. Error bars represent ±1 standard deviation. The number of subjects involved was 27.

sis of variance revealed no statistically significant difference in mean EN contrasting sham versus exposed condition at each of the three stimulus levels ($p > 0.05$).

B. Band-limited TEOAE components

The EN shifts (mean values ±1 standard deviation) in the 33 elementary frequency contributions of TEOAEs recorded before and after a real or sham exposure are presented in Fig. 4 for stimulus level equal to 60 dB peak SPL (panel A), 70 dB peak SPL (panel B), and 80 dB peak SPL (panel C). Maximum mean shifts were, for real and sham exposure, −0.09 and −0.14 mPa² at 60 dB, −2.15 and −1.03 mPa² at 70 dB, and −11.24 and −7.98 mPa² at 80 dB. All these maximum values were found in band 3 (corresponding to the frequency band 1.0–1.1 kHz) except the one after sham exposure at 60 dB, which was measured in band 4 (spanning frequencies from 1.1 to 1.2 kHz). Repeated measures analysis of variance revealed no statistically significant differences ($p > 0.05$) in mean EN due to exposure to EMF for all the band-limited components and for all the three stimulus levels.

Figure 5 shows the mean and standard deviations of LAT variation after a real or sham exposure in the 33 elementary band-limited TEOAE components recorded at 60 dB peak SPL (panel A), 70 dB peak SPL (panel B), and 80 dB peak SPL (panel C) stimulus level. Maximum mean shifts were, for real and sham exposure, respectively, 1.04 and 1.03 ms at 60 dB (bands 2 and 4, respectively, corresponding to the frequency bands 0.9–1.0 kHz and 1.1–1.2 kHz), 0.61 and −0.43 ms at 70 dB (bands 3 and 10, respectively, corresponding to the frequency bands 1.0–1.1 kHz and 1.7–1.8 kHz), and −0.37 and −0.36 ms at 80 dB (bands 8 and 1, respectively, corresponding to the frequency bands 1.5–1.6 kHz and 0.8–0.9 kHz). Repeated measures analysis of variance performed on LAT in all the elementary band-limited TEOAE components revealed that none of the latency shifts was statistically significant ($p > 0.05$) at each of the three eliciting stimulus levels.
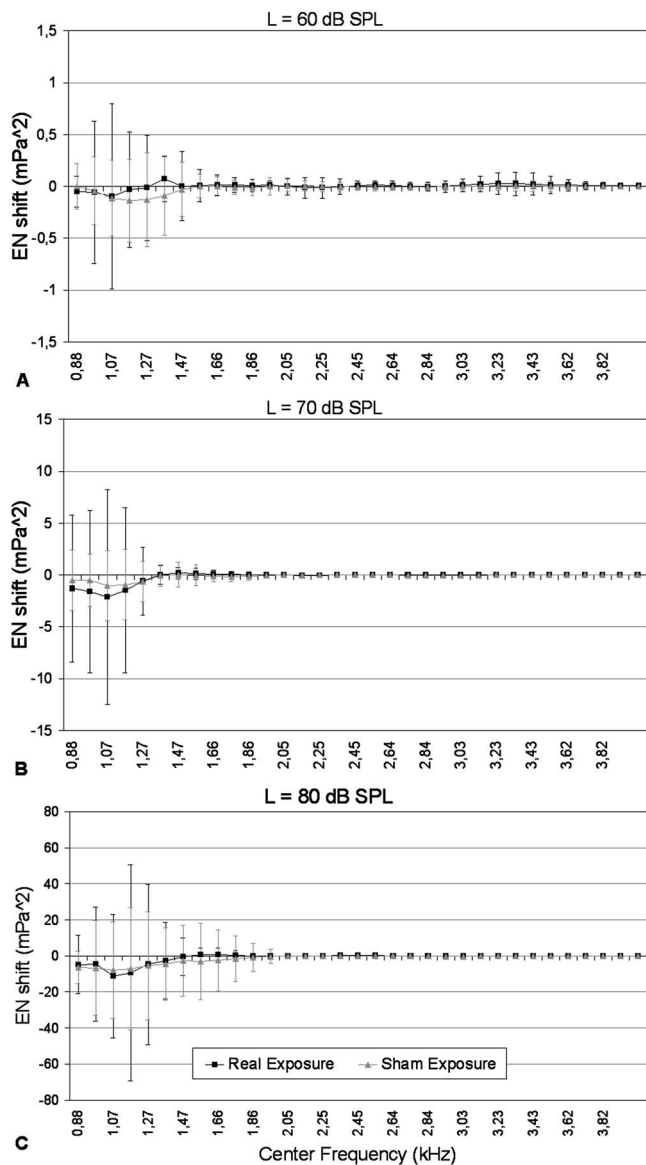
FIG. 4. Energy (EN) shifts in the wavelet elementary band-limited components of TEOAE recorded before and after a real (in black) or sham (in gray) exposure to a GSM mobile phone for the three stimulus levels: 60 dB peak SPL (panel A), 70 dB peak SPL (panel B), and 80 dB peak SPL (panel C). Error bars represent ±1 standard deviation.

FIG. 5. Latency (LAT) shifts in the wavelet elementary band-limited components of TEOAE recorded before and after a real (in black) or sham (in gray) exposure to a GSM mobile phone for the three stimulus levels: 60 dB peak SPL (panel A), 70 dB peak SPL (panel B), and 80 dB peak SPL (panel C). Error bars represent ±1 standard deviation.

IV. CONCLUSIONS

As mentioned above, to date a relatively limited number of studies have addressed the question of possible effects on the hearing system after exposure to a GSM mobile phone and the hearing system and no studies at all have yet explored the effects of EMF nonionizing radiations on energy and latency of TEOAE time-frequency fine structure.

The focus of this paper was to study possible subtle changes in OHC function due to EMFs emitted by a GSM mobile phone that can be investigated by means of a frequency-specific analysis applied to TEOAE recordings. The rationale behind this analysis is that, separating the narrow-band contributions to the evoked response, some insight might be achieved into which elementary components of TEOAEs, if any, may be changing due to exposure.

First of all, the analysis of the broadband TEOAE responses was performed, resulting in no statistically significant effect on TEOAE energy (parameter EN in Fig. 3) due to exposure. The statistically nonsignificant differences observed after real exposure can be considered to lie within the intrinsic variability of the parameter, these shifts being even lower than shifts induced by sham exposure, the latter known to be due to natural oscillations of TEOAE energy among different recordings.

The analysis of the time-frequency fine structure showed no significant effects related to exposure to GSM phones either. No significant effect on energy was found in the band-limited components at any of the three stimulus levels (parameter EN in Fig. 4). The same was true for latency (parameter LAT in Fig. 5). On average, latency changed less than 1 ms between recordings made before and after expo-

sure (real or sham). The sign of the shift was almost balanced between a decrease and an increase in latency after exposure, and all the distributions lay almost symmetrically around zero, confirming that the observed lags and leads were induced by random variations arising by chance.

In conclusion, the results of this study show that there is no measurable effect on TEOAEs immediately after a 10 min exposure at the maximum power of a GSM mobile phone, neither when TEOAEs were analyzed across the entire frequency range nor when analyzed in their time-frequency fine structure. These findings are fully in line with the conclusions of previous studies on animals and humans (Galloni *et al.*, 2005a, 2005b; Parazzini *et al.*, 2005b; Uloziene *et al.*, 2005, Parazzini *et al.*, 2007b).

ACKNOWLEDGMENTS

Arai, N., Enomoto, H., Okabe, S., Yuasa, K., Kamimura, Y., and Ugawa, Y. (**2003**). "Thirty minutes mobile phone use has no short-term adverse effects on central auditory pathways," Clin. Neurophysiol. **114**, 1390–1394.

Bell, S. L., Parazzini, M., Thuroczy, G., Lutman, M., and Ravazzani, P. (**2004**). "Effects of GSM cellular phones on human hearing: Methodological approach and pilot study," *Second IASTED International Conference on Biomedical Engineering (Biomed 2004)*, pp. 519–522.

Borbely, A. A., Huber, R., Graf, T., Fuchs, B., Gallmann, E., and Achermann, P. (**1999**). "Pulsed high-frequency electromagnetic field affects human sleep and sleep electroencephalogram," Neurosci. Lett. **275**, 207–210.

Brownell, W. E., Bader, C. R., Bertrand, D., and De Ribaupierre, Y. (**1985**). "Evoked mechanical responses in isolated cochlear outer hair cells," Science **227**, 194–196.

Burkhardt, M., and Kuster, N. (**2000**). "Appropriate modeling of the ear for compliance testing handheld MTE with SAR safety limits at 900/1800 MHz," IEEE Trans. Microwave Theory Tech. **48**, 1927–1934.

Cheng, J. (**1995**). "Time-frequency analysis of transient evoked otoacoustic emissions via smoothed pseudo Wigner distribution," Scand. Audiol. **24**, 91–96.

Dallos, P. (**1992**). "The active cochlea," J. Neurosci. **12**, 4575–4585.

Frey, A. H. (**1998**). "Headaches from cellular telephones: Are they real and what are the implications?," Environ. Health Perspect. **106**, 101–103.

Fritze, K., Sommer, B., and Schmitz, B. (**1997**). "Effects of global system for mobile communication (GSM) microwave exposure on blood-brain barrier permeability in rat," Acta Neuropathol. (Berl) **94**, 465–470.

Fukai, N., Shyu, J., Driscoll, C., and Kei, J. (**2005**). "Effects of body position on transient evoked otoacoustic emissions: The clinical perspective," Int. J. Audiol. **44**, 8–14.

Galloni, P., Parazzini, M., Piscitelli, M., Pinto, R., Lovisolo, G. A., Tognola, G., Marino, C., and Ravazzani, P. (**2005a**). "Electromagnetic fields from mobile phones do not affect the inner auditory system of Sprague-Dawley rats," Radiat. Res. **164**, 798–804.

Galloni, P., Lovisolo, G. A., Mancini, S., Parazzini, M., Pinto, R., Piscitelli, M., Ravazzani, P., and Marino, C. (**2005b**). "Effects of 900 MHz electromagnetic fields exposure on cochlear cells' functionality in rats: Evaluation of distortion product otoacoustic emissions," Bioelectromagnetics (N.Y.) **26**, 536–547.

Grisanti, G., Parlapiano, C., Tamburello, C. C., Tine', G., and Zanforlin, L. (**1998**). "Cellular phones effects on otoacoustic emissions," IEEE MTT-S Int. Microwave Symp. Dig. 7–12 Jun 1998, **2**, 771–774.

Haarala, C., Aalto, S., Hautzel, H., Julkunen, L., Rinne, J. O., Laine, M., Krause, B., and Hamalainen, H. (**2003**). "Effects of a 902 MHz mobile phone on cerebral blood flow in humans: A PET study," NeuroReport **14**, 2019–2023.

Hamblin, D. L., Wood, A. W., Croft, R. J., and Stough, C. (**2004**). "Examining the effects of electromagnetic fields emitted by GSM mobile phones on human event-related potentials and performance during an auditory task," Clin. Neurophysiol. **115**, 171–178.

Janssen, T., Boege, P., Oestreicher, E., and Arnold, W. (**2000**). "Tinnitus and $2f_1$-f_2 distortion product otoacoustic emissions following salicylate overdose," J. Acoust. Soc. Am. **107**, 1790–1792.

Janssen, T., Boege, P., von Mikusch-Buchberg, J., and Raczek, J. (**2005**). "Investigation of potential effects of cellular phones on human auditory function by means of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **117**, 1241–1247.

Jedrzejczak, W. W., Blinowska, K. J., Konopka, W., Grzanka, A., and Durka, P. J. (**2004**). "Identification of otoacoustic emissions components by means of adaptive approximations," J. Acoust. Soc. Am. **115**, 2148–2158.

Kellenyi, L., Thurockzy, G., Faludy, B., and Lenard, L. (**1999**). "Effects of mobile GSM radiotelephone exposure on the auditory brainstem response (ABR)," Neurobiology **7**, 79–81.

Kemp, D. T. (**1979**). "Evidence of mechanical non-linearity and frequency selective wave amplification in the cochlea," Arch. Otolaryngol. **224**, 37–45.

Kizilay, A., Ozturan, O., Erdem, T., Kalcioglu, M. T., and Miman, M. C. (**2003**). "Effects of chronic exposure of electromagnetic fields from mobile phones on hearing in rats," Auris Nasus Larynx **30**, 239–245.

Kummer, P., Janssen, T., and Arnold, W. (**1998**). "The level and growth behavior of the $2f_1$-f_2 distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss," J. Acoust. Soc. Am. **103**, 3431–3444.

Lonsbury-Martin, B. L., and Martin, G. K. (**1990**). "The clinical utility of distortion-product otoacoustic emissions," Ear Hear. **11**, 144–154.

Mallat, S. G. (**1989**). "A theory for multiresolution signal decomposition: The wavelet representation," IEEE Trans. Pattern Anal. Mach. Intell. **11**, 674–693.

Marino, C., Cristalli, G., Galloni, P., Pasqualetti, P., Piscitelli, M., and Lovisolo, G. A. (**2000**). "Effects of microwaves (900 MHz) on the cochlear receptor: Exposure systems and preliminary results," Radiat. Environ. Biophys. **39**, 131–136.

Marozas, V., Janusauskas, A., Lukosevicius, A., and Sornmo, L. (**2006**). "Multiscale detection of transient evoked otoacoustic emissions," IEEE Trans. Biomed. Eng. **53**, 1586–1593.

McFadden, D., and Plattsmier, H. S. (**1984**). "Aspirin abolishes spontaneous otoacoustic emissions," J. Acoust. Soc. Am. **76**, 443–448.

Moleti, A., Sisto, R., Tognola, G., Parazzini, M., Ravazzani, P., and Grandori, F. (**2005**). "Otoacoustic emission latency, cochlear tuning, and hearing functionality in neonates," J. Acoust. Soc. Am. **118**, 1576–1584.

Morand, N., Khalfa, S., Ravazzani, P., Tognola, G., Grandori, F., Durrant, J. D., Collet, L., and Veuillet, E. (**2000**). "Frequency and temporal analysis of contralateral acoustic stimulation on evoked otoacoustic emissions in humans," Hear. Res. **145**, 52–58.

National Radiological Protection Board, Chilton, UK. (**2003**). "Health effects from radio frequency electromagnetic fields—Report of an independent advisory group on non-ionizing radiation," National Radiological Protection Board, Documents of the NRPB **14**, No. 2.

O'Brien, A. (**1994**). "Temperature dependency of the frequency and level of a spontaneous otoacoustic emission during fever," Audiology **28**, 281–290.

Oftedal, G., Wilen, J., Sandstrom, M., and Mild, K. H. (**2000**). "Symptoms experienced in connection with mobile phone use," Rev. Asoc Prof. Hosp. Nac. Odontol. **50**, 237–245.

Ozturan, O., Erdem, T., Miman, M. C., Kalcioglu, M. T., and Oncel, S. (**2002**). "Effects of the electromagnetic field of mobile telephones on hearing," Arch. Otolaryngol. **122**, 289–293.

Parazzini, M., Galloni, P., Piscitelli, M., Pinto, R., Lovisolo, G. A., Tognola, G., Ravazzani, P., and Marino, C. (**2007a**). "Possible combined effects of 900 MHz CW electromagnetic fields and gentamicin on the auditory system of rats," Radiat. Res. **167**, 600–605.

Parazzini, M., Brazzale, A. R., Paglialonga, A., Tognola, G., Collet, L., Moulin, A., Lutman, M. E., Bell, S. L., Thomas, N. A., Uloziene, I., Uloza, V., Thuroczy, G., Tavartkiladze, G., Tsalighopoulos, M., Kyriafinis, G., and Ravazzani, P. (**2007b**). "Effects of GSM cellular phones on human hearing: The European project 'GUARD'," Radiat. Res. **167**.

Parazzini, M., Hall, A. J., Lutman, M. E., and Kapadia, S. (**2005a**). "Effect of aspirin on phase gradient of 2F1-F2 distortion product otoacoustic emissions," Hear. Res. **205**, 44–52.

Parazzini, M., Bell, S., Thuroczy, G., Molnar, F., Tognola, G., Lutman, M. E., and Ravazzani, P. (**2005b**). "Influence on the mechanisms of generation of distortion product otoacoustic emissions of mobile phone exposure," Hear. Res. **208**, 68–78.

Preece, A. W., Iwi, G., and Davies-Smith, A. (**1999**). "Effect of a 915 MHz simulated mobile phone signal on cognitive function in man," Int. J. Radiat. Biol. **75**, 447–456.

Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2067.

Ravazzani, P., Collet, L., Lutman, M. E., Tavartkiladze, G., Thuroczy, G., Tsalighopoulos, M., Uloza, V., Uloziene, I., Bell, S. L., Moulin, A., Parazzini, M., Thomas, N. A., and Tognola, G. (**2005**). "Effects of GSM cellular phones on human hearing: The European Project GUARD," *Association for Research in Otolaryngology, 28th ARO MidWinter Meeting abstracts*, p. 60.

Skellett, R. A., Crist, J. R., Fallon, M., and Bobbin, R. P. (**1996**). "Chronic low-level noise exposure alters distortion product otoacoustic emissions," Hear. Res. **98**, 68–76.

Tognola, G., Grandori, F., and Ravazzani, P. (**1997**). "Time-frequency distributions of click-evoked otoacoustic emissions," Hear. Res. **106**, 112–122.

Tognola, G., Grandori, F., and Ravazzani, P. (**1998**). "Wavelet analysis of click-evoked otoacoustic emissions," IEEE Trans. Biomed. Eng. **45**, 686–697.

Uloziene, I., Uloza, V., Gradauskiene, E., and Saferis, V. (**2005**). "Assessment of potential effects of the electromagnetic fields of mobile phones on hearing," BMC Public Health **5**, 39.

Wit, H. P., van Dijk, P., and Avan, P. (**1994**). "Wavelet analysis of real and synthesized click evoked otoacoustic emissions," Hear. Res. **73**, 141–147.

Yang, L. P., Young, S. T., and Kuo, T. S. (**2003**). "Combination of derived non-linear and linear methods to increase the reproducibility of transiently evoked otoacoustic emissions," Int. J. Audiol. **42**, 331–338.

Yang, X., Wang, K., and Shamma, S. A. (**1992**). "Auditory representations of acoustic signals," IEEE Trans. Inf. Theory **38**, 824–839.

Zheng, J., Shen, W., He, D. Z., Long, K. B., Madison, L. D., and Dallos, P. (**2000**). "Prestin is the motor protein of cochlear outer hair cells," Nature (London) **405**, 149–155.

Transient evoked otoacoustic emission latency and cochlear tuning at different stimulus levels

Renata Sisto[a)]

Dipartimento Igiene del Lavoro, ISPESL, Via Fontana Candida, 1, 00040 Monte Porzio Catone, Roma, Italy

Arturo Moleti[b)]

Dipartimento di Fisica, Università di Roma "Tor Vergata," Via della Ricerca Scientifica, 1, 00133 Roma, Italy

Cochlear latency has been evaluated in young adults by time-frequency analysis of transient evoked otoacoustic emissions recorded using the nonlinear acquisition mode at different levels of the click stimulus. Objective, even if model-dependent, estimates of cochlear tuning have been obtained from the otoacoustic latency estimates. Transmission-line cochlear models predict that the transient-evoked otoacoustic emission latency is dependent on the stimulus level, because the bandwidth of the cochlear filter (tuning) depends on the local cochlear excitation level due to nonlinear damping. The results of this study confirm the increase of tuning with increasing frequency and show clearly the decrease of latency and tuning with increasing stimulus level. This decrease is consistent with the expected relation between the slowing down of the traveling wave near the tonotopic place and the cochlear excitation amplitude predicted by cochlear models including nonlinear damping. More specifically, these results support the models in which nonlinear damping consists of a quadratic term and a constant positive term. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2769981]

I. INTRODUCTION

Otoacoustic emissions (OAE) are acoustical signals recorded in the ear canal, which are generated in the cochlea, either as a response evoked by an acoustical stimulation (evoked OAEs, or EOAEs) or in the absence of any stimulation (spontaneous OAEs, or SOAEs). In physiological acoustics, several different OAE techniques (Probst *et al.* 1991) have been used in connection with theoretical cochlear models to accurately study the response of the inner ear. It is commonly accepted that OAEs are a by-product of the cochlear active amplification mechanisms.

Experimentally, the latency of transient-evoked OAEs (TEOAEs) is a function of both the frequency of the emission and the stimulus level. The relation among TEOAE latency, frequency, and stimulus level has been evaluated using tone bursts (Norton and Neely, 1987; Neely *et al.*, 1988; Jedrzeczjak *et al.*, 2004) and clicks (Wit *et al.*, 1994; Tognola *et al.*, 1997; Sisto and Moleti, 2002; Moleti and Sisto, 2003; Jedrzeczjak *et al.*, 2004; Moleti *et al.*, 2005). In the more recent studies, the TEOAE latency was evaluated using time-frequency analysis techniques, based on wavelet and matching pursuit (MP) algorithms. MP algorithms use an iterative procedure to represent the TEOAE wave form as a superposition of "atoms," which are elementary oscillating functions, localized in the time-frequency domain (Mallat and Zhang, 1993). The first atom is chosen from a redundant

"dictionary" of functions, by maximizing its inner product with the original TEOAE wave form. The first "residual" is obtained by subtracting from the original wave form its projection along the selected atom, and the procedure is iteratively repeated on the residuals until convergence is reached. Typically, 10 to 20 atoms are sufficient to accurately reconstruct the original TEOAE wave form (Jedrzeczjak *et al.*, 2004).

Time-domain estimates of the OAE latency have also been obtained measuring the onset time of distortion product OAEs (DPOAEs), for both the $2f_1 - f_2$ and the $2f_2 - f_1$ emissions (Whitehead *et al.*, 1996; Martin *et al.*, 1998). Unfortunately, the DPOAE onset delays cannot be easily compared with the TEOAE latencies, because there are two DPOAE sources and because the backward DP wave has a frequency that is different from those of the primary tones.

Other studies (e.g., Prieve *et al.*, 1996) evaluated the TEOAE group delay measuring the slope of the phase-frequency relation. This characteristic time, more accurately named phase-gradient delay, has been extensively analyzed by several authors in the case of stimulus frequency OAEs (SFOAEs) (Shera and Guinan, 2003) and DPOAEs (Moulin and Kemp, 1996), and also compared to electrophysiological delay measurements (Schoonhoven *et al.*, 2001). It is related to the cochlear transmission delay in a model-dependent way, according to the wave-fixed or place-fixed nature of the OAE generation mechanism, and we will not discuss this relation in this work.

The frequency dependence of the OAE latency is predicted by cochlear models, as a consequence of the relation

---

[a)]Electronic mail: renata.sisto@ispesl.it
[b)]Electronic mail: moleti@roma2.infn.it

between the frequency of the emission and the position along the cochlea of its generation place. For TEOAEs, the source is assumed to be close to the resonant place, whose position depends on frequency according to the Greenwood tonotopic map (Greenwood, 1990). The study of the latency-frequency relation permits one to test the theoretical models for the transmission of the acoustic signal (traveling wave, or TW) along the basilar membrane (BM).

In the models based on the transmission-line formalism (e.g., Furst and Lapid, 1988; Talmadge et al., 1998), the cochlea is described as an equivalent electrical transmission line, which is locally resonant at a characteristic frequency, according to the Greenwood map. The transmission velocity along the BM for a given frequency component of the stimulus and of the OAE response is the group velocity, associated with the local relation between the wave vector and the frequency. Typically, a complete solution of the cochlear model equations requires numerical techniques, but analytical approximate results can be obtained, making simplifying assumptions. One of these assumptions is the scale invariance symmetry hypothesis, which postulates that the physics of the cochlea is independent of the frequency scale. This symmetry is only approximately verified in the real cochlea. It can be easily demonstrated that in fully scale-invariant models, an inverse proportionality relation is expected between latency and frequency (Talmadge et al., 1998). Experimentally, this relation is described by a much slower power law, with an exponent between −0.4 and −0.7 (e.g., Tognola et al., 1997; Jedrzeczjak et al., 2004; Moleti et al., 2005). The breaking of the scaling symmetry is also suggested by other phenomenological evidence: (1) At low frequency the Greenwood map significantly departs from a logarithmic relation, introducing an explicit frequency scale, and (2) psychoacoustical and OAE estimates of cochlear tuning (Shera et al., 2002) suggest that the cochlear quality factor Q is not a constant, as it must be in SI models.

Under a few model-dependent assumptions, the experimental relation between latency and frequency has permitted one to get objective estimates of the frequency dependence of the cochlear quality factor Q, which is defined as the characteristic frequency divided by the bandwidth of the resonance (Shera et al., 2002; Shera and Guinan, 2003; Moleti and Sisto, 2003). The underlying hypothesis is that the slowing-down of each frequency component of the traveling wave near its resonant place is dependent on the bandwidth of the local cochlear active filter. As the bandwidth is a function of the local BM excitation level, measurements of the TEOAE latency as a function of the stimulus level may provide information about the nonlinear nature of the cochlear active filter mechanism. This mechanism is associated with the feedback mediated by the outer hair cells, which is responsible for the remarkable tuning and sensitivity performances of the healthy inner ear.

In this study, the TEOAE latency will be evaluated as a function of the stimulus level, to get information on the dependence of cochlear tuning on the stimulus level, and to compare the results with the predictions of transmission-line cochlear models. The main purpose of these quality factor estimates is not to provide an alternative technique for measuring cochlear tuning, but rather to link different findings from OAE experiments with cochlear models to get a more coherent framework for understanding the inner ear mechanisms.

II. METHODS

TEOAE data from 10 young subjects have been analyzed, recorded with the ILO 292 Echoport system (Otodynamics, Ltd.), using the nonlinear acquisition paradigm. In the nonlinear acquisition mode, three identical click stimuli are followed by a fourth click of triple amplitude and opposite polarity, and the average of these four 20-ms-long responses is defined as the "nonlinear" response. Its main purpose is to cancel the so-called linear ringing artifact, along with any other component of the response that is proportional to the stimulus amplitude. Two alternate averages, each of 260 such wave forms, are recorded, bandpass filtered (from 600 Hz to 6 kHz), and windowed (onset at 2.5 ms and flattop from 5 ms). The stimulus level was varied between 60 and 90 dB pSPL in 5 dB steps. The FFT of each wave form was also computed and its modulus averaged over seven frequency bands, from 1 to 4 kHz, to get an estimate of the OAE response in each band.

A time-frequency analysis technique based on the continuous wavelet transform (CWT) was used to estimate the OAE latency as a function of frequency (Tognola et al., 1997; Moleti et al., 2005). Wavelet analysis projects the signal on a basis of elementary wavelet functions that are dilated and time-translated versions of a "mother wavelet" $\psi(t)$. The wavelet transform of a signal $x(t)$ at the time τ and frequency f is defined by the inner product:

$$\text{WT}(\tau,f) = \int x(t)\left(\frac{f}{f_0}\right)^{-3/2} \psi^*\left(\frac{f}{f_0}(t-\tau)\right)dt, \qquad (1)$$

where the wavelet inside the integral is a function centered in the time domain at the time τ, and, in the frequency domain, at the frequency f.

Equation (1) defines a set of wavelet coefficients that may be seen as functions of time, whose characteristic frequencies are distributed along the frequency axis. Integrating these coefficients over 500 Hz frequency bands, seven coefficients $\text{WT}(\tau,f_i)$ relative to bands centered between 1 and 4 kHz have been obtained. Outside this frequency interval, the estimates of the relation among latency, frequency, and stimulus level are affected by systematic errors associated with the acquisition window and the stimulus spectral density.

The spectral latency relative to the ith frequency band may be defined as the time $t_{\max}(f_i)$ for which the coefficient of the ith band, $\text{WT}(\tau,f_i)$, reaches its maximum absolute value. Using this definition it is possible to obtain a statistical distribution of the individual spectral latency in a sample of ears, by evaluating $t_{\max}(f_i)$ for each ear. Experimentally, it turns out that this distribution is not Gaussian and strongly asymmetrical, particularly at frequencies higher than 2 kHz, characterized by long-delay tails. Therefore, the average, called τ^* in the following, and the variance are not sufficient to describe the properties of the distribution. Moreover, it has

R. Sisto and A. Moleti: Otoacoustic emission latency and tuning

been noted that the unreasonably long-latency values that affect the high-frequency average latencies occur more frequently when the signal to noise ratio (SNR) is particularly low (Moleti et al., 2005), thus systematic errors may affect the study of the dependence of latency on the stimulus level.

As an alternative, it is possible to average over the ears the normalized wavelet coefficients, defining as average latency of the population the time for which the envelope of the average wavelet coefficient reaches its maximum absolute value. The envelope, computed applying a moving time-average and a polynomial fit to the average wavelet coefficient, was used to find the time of the maximum, minimizing the uncertainty due to the oscillations of the wavelet coefficients. This definition provides a single latency value, called τ in the following, and the temporal width of the peak of the average wavelet coefficient may provide a rough estimate of the uncertainty due to intersubject variability and noise. This procedure yields more stable results, and it is less sensitive to the systematic errors occurring at low SNR levels (Moleti et al., 2005), and to the presence of SOAEs, whose latency/frequency relation is also anomalous.

In Sisto and Moleti (2002) and in Moleti and Sisto (2003), a line identification analysis method had also been used, in which the band wavelet coefficient was compared to the spectrum of the TEOAE wave form to identify the latency of a few frequency components. This procedure may avoid some systematic errors associated with attributing to each wavelet component the mean (or geometric mean) frequency of the correspondent band, but it is rather subjective and time-consuming, and cannot be reasonably proposed for automatic TEOAE latency estimates.

While the time-frequency latency estimate is a rather direct one, to get information on cochlear tuning from latency measurements it is necessary to make some theoretical assumptions. In cochlear models tuning is related indeed in a model-dependent way to the decrease of the transmission velocity of each frequency component of the TW near its resonant place. Due to this dependence on cochlear model assumptions, we will discuss this procedure in Sec. III.

III. MODEL

The TEOAE latency is essentially due to the roundtrip transmission time of the acoustic wave along the BM from the base to the tonotopic place. Due to its tonotopic structure, the cochlea produces temporal dispersion of the frequency components of the click stimulus, resulting in the particular time-frequency shape of the TEOAE response. The form of the Greenwood map implies that the higher frequency components of the stimulus travel a shorter path to reach their tonotopic place, where the corresponding component of the retrograde wave (the OAE) is generated. As a consequence, the OAE latency is a decreasing function of frequency. Another important issue is related to the slowing-down of each frequency component of the TW near its resonant place, which is a function of the quality factor of the resonance. For a quantitative interpretation of the latency-frequency relation it is necessary to make explicit reference to a particular cochlear model. An important class of models is based on the

formalism of the equivalent electrical transmission line (Furst and Lapid, 1988; Talmadge et al., 1998; Shera et al., 2005). In this study we will use a simple one-dimensional model (Moleti and Sisto, 2003), in which the line is locally resonant at the frequency $\omega(x)$ at the place x along the BM, while the other mechanical and geometrical parameters of the cochlea (cross-sectional area, membrane surface density, etc.) are included in a single dimensional parameter k_0. A local relation is obtained between the wave vector k, the angular frequency $\omega = 2\pi f$, and the longitudinal position x:

$$k(\omega,x) = \frac{\omega k_0}{\sqrt{\Delta(x,\omega)}} = \frac{\omega k_0}{\sqrt{\omega^2(x) - \omega^2 + i\omega\Gamma}}. \quad (2)$$

The transmission velocity of the wave packet of frequency ω may be computed from Eq. (2), as a function of x, as the group velocity v_g. The cochlear latency $\tau(\omega)$ may be estimated for each frequency as the integral of the inverse of the traveling wave velocity $v_g(\omega,x)$ extended from the base to the tonotopic place, if a wave-fixed model is assumed in which each OAE component is generated exactly at its resonant place, where the group velocity inverts itself:

$$\tau(\omega) = 2\int_0^{x(\omega)} \frac{dx}{v_g(\omega,x)} = 2\int_0^{x(\omega)} \frac{\partial \operatorname{Re}(k(\omega,x))}{\partial \omega} dx. \quad (3)$$

Using the scale invariance hypothesis for the part of the integral far from the resonance (Moleti and Sisto, 2003), it is possible to separate a contribution due to the resonance [within $\Delta x = 2\Gamma/\omega k_\omega$ from $x(\omega)$], which is independent of Q and decreases with frequency as ω^{-1}, while the contribution of the part of the path far from the resonance [up to $x(\omega) - \Delta x$], is a function of Q:

$$\tau_{\mathrm{res}}(\omega) = 2\int_{x(\omega)-\Delta x}^{x(\omega)} dx \frac{\partial k}{\partial \omega} \cong \sqrt{2}\frac{k_0}{\omega k_\omega}, \quad (4)$$

$$\tau_{\mathrm{far}}(\omega) \cong 2\int_0^{x(\omega)-\Delta x} dx \frac{\partial k}{\partial x}\frac{\partial x}{\partial \omega}$$

$$= \frac{2(k(x(\omega)-\Delta x) - k(0))}{\omega k_\omega} = \frac{k_0}{\omega k_\omega}\left(\alpha\sqrt{Q} - \frac{2\omega}{\omega_0}\right), \quad (5)$$

where $\alpha \approx 0.9$ is a dimensionless constant and $\omega_0 = 2\pi \times 20.655$ kHz. These relations permit one to obtain an objective estimate of cochlear quality factor, based on experimental TEOAE latency data $\tau(\omega)$:

$$Q(\omega) = \frac{1}{\alpha^2}\left(\frac{\omega k_\omega}{k_0}(\tau(\omega) - \tau_{\mathrm{res}}(\omega) - \tau_{\mathrm{nc}}) + \frac{2\omega}{\omega_0}\right)^2, \quad (6)$$

where τ_{nc} is a noncochlear delay, evaluated of order 0.5 ms (Sisto and Moleti, 2002). According to the coherent reflection filtering theory (Talmadge et al., 1998; Shera et al., 2005), place-fixed generation of backward waves occurs instead within a cochlear region that is slightly basal to the resonant place (Talmadge et al., 1998; Shera et al., 2005), due to the presence of randomly distributed mechanical irregularities ("roughness") that act coherently as backscattering centers. To account for this shift of the source position, it

is more correct to neglect the contribution τ_{res} in the estimate of the quality factor and to extend the integral of Eq. (5) to $x(\omega) - \Delta x/2$, where the activity pattern approximately reaches its maximum, obtaining

$$Q(\omega) = \frac{1}{\beta^2}\left(\frac{\omega k_\omega}{k_0}(\tau(\omega) - \tau_{\text{nc}}) + \frac{2\omega}{\omega_0}\right)^2, \qquad (7)$$

where $\beta \approx 1.2$ is another dimensionless constant.

Equations (6) and (7) permit one to approximately evaluate the cochlear quality factor from experimental latency estimates, assuming, respectively, wave-fixed emission from the resonant cochlear place, and place-fixed reflection from a cochlear region centered near the peak of the activity pattern. To be more accurate, a fully consistent place-fixed OAE generation model should be modified to include time-delayed stiffness terms (Zweig and Shera, 1995; Talmadge *et al.*, 1998; Shera *et al.*, 2005), and the integrand of Eq. (3) should be modified accordingly. These terms act as effective damping and antidamping terms, and their inclusion is very important to match the experimental data on the amplitude and phase of the BM transfer function, as first demonstrated by Zweig (1991). As regards the computation of the TEOAE latency only, the inclusion of fast and slow time-delayed stiffness terms only affects the slowing down of the traveling wave by modifying the functional form of the real part of the wave vector near the resonance. This effect may be schematized by simply introducing an effective Γ, so, in the following, we will keep using Eq. (2), with the agreement that the quality factor of Eq. (7) is associated with this effective Γ.

In transmission-line models, cochlear nonlinearity is described introducing a damping function, the real part of the transverse line impedance, which is a nonlinear function of the local BM transverse displacement $\xi(x)$. This term usually includes a nonlinear damping function proportional to the square of the local displacement, as in the Van der Pol oscillator. Other oscillator equations have been proposed with different forms of damping and antidamping terms (e.g., van Hengel *et al.*, 1996; Sisto and Moleti, 1999). In cochlear models using time-delayed stiffness terms (Talmadge *et al.*, 1998), nonlinear damping may be explicitly introduced with a Van der Pol damping term, but also effective linear damping and antidamping are actually present, within the slow and fast delayed-stiffness terms.

In the nonlinear damping function suggested by Sisto and Moleti (1999):

$$\Gamma(\xi) = A\xi^2 + B - \frac{C}{\langle \xi^2 \rangle}, \qquad (8)$$

an increase of the damping coefficient with increasing stimulus level is predicted at high stimulus levels, as the first nonlinear term of Eq. (8) becomes the dominant one, while the last antidamping term is effective only at the lowest levels of excitation, allowing for the existence of limit-cycle oscillations associated with SOAEs. Of course, a quantitative comparison with the theory requires evaluating the dependence of Γ on the excitation amplitude ξ, and not on the stimulus level. Unfortunately, this information is not directly available. It is possible to make some additional assumptions,
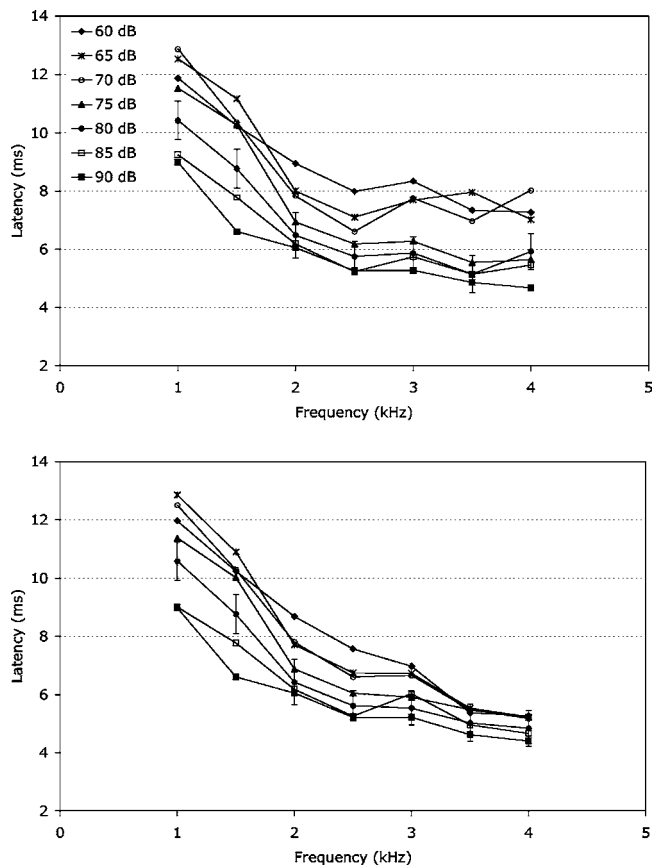
FIG. 1. Average TEOAE latency, as a function of frequency, at different stimulus levels, obtained as the mean (top panel) of the distribution of the individual latencies. Due to the asymmetry of the distributions, the high-frequency mean latencies are overestimated. A much better result is obtained (bottom panel) if the individual latencies from frequency bands with TEOAE SNR < 1 are excluded from each average.

e.g., that the excitation amplitude at a given cochlear place be proportional, for a given ear, to the amplitude of the corresponding otoacoustic emission. This is not necessarily true for all OAE generation mechanisms, for example it would not be true for nonlinear distortion, and we will put it forward in Sec. IV mainly as a hypothesis, useful to graphically present the results in a form closer to the theoretical interpretation point of view.

IV. RESULTS AND DISCUSSION

A. TEOAE latency at different stimulus levels

At frequencies lower than 1 kHz, the short acquisition interval (20 ms) may cause a systematic underestimate of the average latency, while the opposite problem could arise at frequencies higher than 4 kHz, due to the window onset between 2.5 and 5 ms. For this reason, we have chosen to present only the results for the seven frequency bands between 1 and 4 kHz. In the top panel of Fig. 1 we show the experimental estimates of the average spectral latency $\tau^*(f)$, defined as the ear average of the individual latencies, for seven different stimulus levels, from 60 to 90 dB pSPL. Standard errors are shown (for the 80 dB curve only, for clarity), to give a hint of the statistical significance of the observed differences. Although the qualitative behavior is

R. Sisto and A. Moleti: Otoacoustic emission latency and tuning
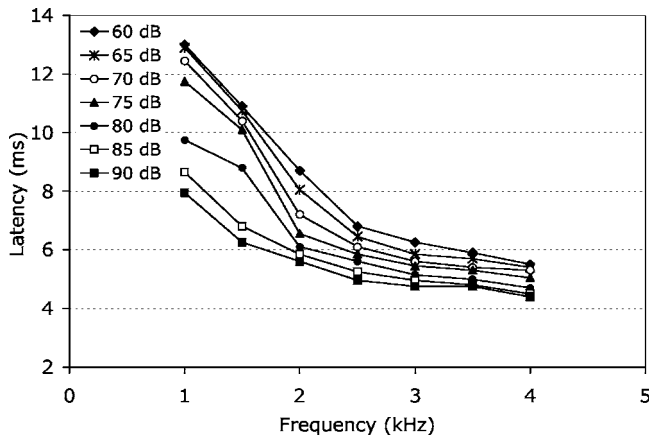
FIG. 2. TEOAE latency, as a function of frequency and stimulus level, obtained by averaging the individual normalized wavelet coefficients. A regular decrease of latency with increasing frequency and stimulus level is clearly visible.

TABLE I. Best fit parameters of Eq. (9). The parameters a and b are obtained by fitting the latency data from 1 to 4 kHz, while $a^{\#}$ and $b^{\#}$ are relative to a fit to all the band latencies from 500 Hz to 5 kHz. Systematic errors at the lowest and highest frequencies, due to the acquisition window and stimulus, result in a large difference between the fit parameters, particularly at low stimulus levels.

Stimulus (dB)	a (ms)	b	$a^{\#}$ (ms)	$b^{\#}$
60	13.4	−0.66	10.8	−0.44
65	13.1	−0.68	10.4	−0.45
70	12.4	−0.68	10.2	−0.48
75	11.7	−0.66	9.6	−0.45
80	9.9	−0.57	8.8	−0.45
85	8.3	−0.46	8.0	−0.40
90	7.6	−0.41	7.3	−0.35

that theoretically predicted, i.e., latency decreasing with increasing frequency and stimulus level, the behavior is rather irregular, particularly at the lowest stimulus levels, where the SNR is typically lower. By examining the distributions of the individual latencies it is possible to note that they are not Gaussian, with significant tails at long latency that seriously affect the estimate of the mean of the distribution. In these cases, one could wonder if the average is still the single parameter that best describes the whole distribution. Anomalously long high-frequency latency values are presumably due to noise, as confirmed by the fact that they are more frequent at low stimulus (and SNR) levels. As already noted (Moleti *et al.*, 2005), this fact introduces a systematic bias in the analysis of the dependence of latency on stimulus level and frequency. A selection based on the TEOAE SNR may be adopted (Moleti *et al.*, 2005), at the risk of introducing another bias in the interpretation of the results, because the level of the response is also correlated to latency and tuning. Moreover, this selection rule would not exclude data dominated by the contribution of SOAEs, whose latency-frequency relation is also anomalous. Selecting only the latency data from the frequency bands with SNR > 1 (70% of the total), the result shown in the bottom panel of Fig. 1 is obtained, with significantly shorter high-frequency latencies. It can be noticed that the relation among latency, frequency, and stimulus level is still rather irregular.

The above mentioned results suggest using the alternative definition of average latency, as the time τ at which the average of the normalized wavelet coefficients reaches its maximum absolute value, which is less sensitive to systematic errors. Indeed, a small fraction of noisy wave forms may yield unreasonable individual latency values that seriously affect the mean of the individual latency distribution, while the time of the maximum of the average wavelet coefficient is much more stable, as long as the fraction of noisy data is small. As shown in Fig. 2, the latency computed from the average wavelet coefficients is indeed a much more regular and monotonic function of both frequency and stimulus level.

In Table I the best fit parameters for the curves of Fig. 2 are listed, having used a power law of the form:

$$\tau = af^b \qquad (9)$$

with f in kilohertz. In the last two columns of Table I, the fit parameters $a^{\#}$ and $b^{\#}$ are also shown, which are obtained by also including in the fit lower-frequency (500 Hz) and higher-frequency (4.5 and 5 kHz) latency estimates. As discussed earlier, these estimates are probably affected by systematic errors associated with the acquisition window and the stimulus spectral shape, and their inclusion leads one to underestimate the slope of the latency curve. These results are shown here to appreciate the size of the systematic errors that can be made by neglecting this problem, and to allow comparison with other studies.

Figure 2 and Table I not only show that the latency decreases with increasing stimulus level through the decrease of the multiplicative coefficient a, but also that the slope of the power law fit to TEOAE latency is significantly steeper at lower stimulus levels. The second result is quite new and should be discussed. Tognola *et al.* (1997), using click stimuli between 47 and 68 dB, found slopes between −0.40 and −0.48, with no clear dependence on the stimulus level. A similar result had also been reported by Neely *et al.* (1988) for the auditory brainstem response (ABR) forward delay, in a study using tone bursts, which also showed that the OAE latency was approximately twice the ABR forward delay. They measured a slope of −0.41 at all stimulus levels for the ABR latency, and assumed that the dependence on the stimulus level was in a multiplying factor only, while their OAE latency estimates were affected by large uncertainties.

As regards the OAE latency only, the above-mentioned systematic errors in the estimate of the lowest- and highest-frequency latencies due the acquisition window could be the cause of the discrepancy between our results and Tognola *et al.* (1997). Indeed, if we included in the fit also the 500 Hz, 4.5 kHz, and 5 kHz latency estimates, we would get the fit parameters $a^{\#}$ and $b^{\#}$ shown in the last columns of Table I, in rather good agreement with those found by Tognola *et al.* (1997). As regards the comparison with the slope of the ABR forward latency, it should be noted that the fit parameters are dependent on the assumed value of the neural delay offset, which in Neely *et al.* (1998) was assumed to be 5 ms, inde-

pendent of the stimulus level. In the case of OAEs, there is also a noncochlear offset delay τ_{nc} to be considered, which has not been included in the fit at this stage, to allow a clearer comparison with other OAE studies. A slightly different value of these offsets would alter the balance between the two fit coefficients.

Sisto and Moleti (2002) found $a=9.2$ ms and $b=-0.37$ for $77-80$ dB clicks, using octave band wavelet analysis, and $a=10.2$ ms and $b=-0.68$, using a line identification method and including a noncochlear offset of 0.5 ms. Assuming no offset the fit to the same data would have given $a=10.7$ ms and $b=-0.63$, which would be compatible with the results of the present study at stimulus levels of $75-80$ dB. The lower slope obtained with octave band analysis was attributed to systematic errors associated with attributing the central frequency of the band to a signal dominated by spectral components near the inner boundaries of the extreme bands (Sisto and Moleti, 2002).

Using $65-68$ dB clicks, Jedrzejczak *et al.* (2004) found, with a matching pursuit technique, the latency fit parameters $a=10.8$ and $b=-0.56$, while they found $a=11.5$ and $b=-0.62$ using tone bursts. These fit parameters are somewhere in between those of the two fits shown in Table I, which is not surprising, considering that their analysis includes atoms from a frequency range whose width is intermediate between the two choices.

The DPOAE onset latency (Whitehead *et al.*, 1996; Martin *et al.*, 1998) is typically shorter than the TEOAE latency measured both in this and in the other cited studies, for both the $2f_1-f_2$ and the $2f_2-f_1$ emissions. As already pointed out by Whitehead *et al.* (1996), it is expected that the $2f_1-f_2$ emission latency should be shorter than the TEOAE latency at frequency f_2, because the backward wave is emitted from the cochlear place resonant at f_2, $x(f_2)$, which is more basal than its own resonant place $x(2f_1-f_2)$. As a consequence, the group velocity of the $2f_1-f_2$ retrograde wave generated in $x(f_2)$ would be higher than that of the f_2 wave, along the same path. The complex nature of the DPOAE generation mechanisms makes it difficult to give a quantitative interpretation to this comparison, which is further complicated by the significantly shorter onset delays measured for the $2f_2-f_1$ emissions (Martin *et al.*, 1998) in agreement with phase-gradient delay studies (e.g., Wable *et al.*, 1996). This would suggest either hypothesizing a $2f_2-f_1$ source more basal than the f_2 place, or assuming most of the $2f_1-f_2$ emission be generated from its "second source" at the $2f_1-f_2$ place, which is not expected at the rather high primary levels (75 dB) used in that study.

The TEOAE group delay measurements by Prieve *et al.* (1996) show significantly higher latencies, particularly at the highest frequencies. It should be stressed that the relation between the group delay and the transmission delay measured by time-frequency analysis is not obvious, and it is crucially dependent on the place-fixed or wave-fixed nature of the assumed OAE generation mechanism. Here, we will not go further into this discussion.
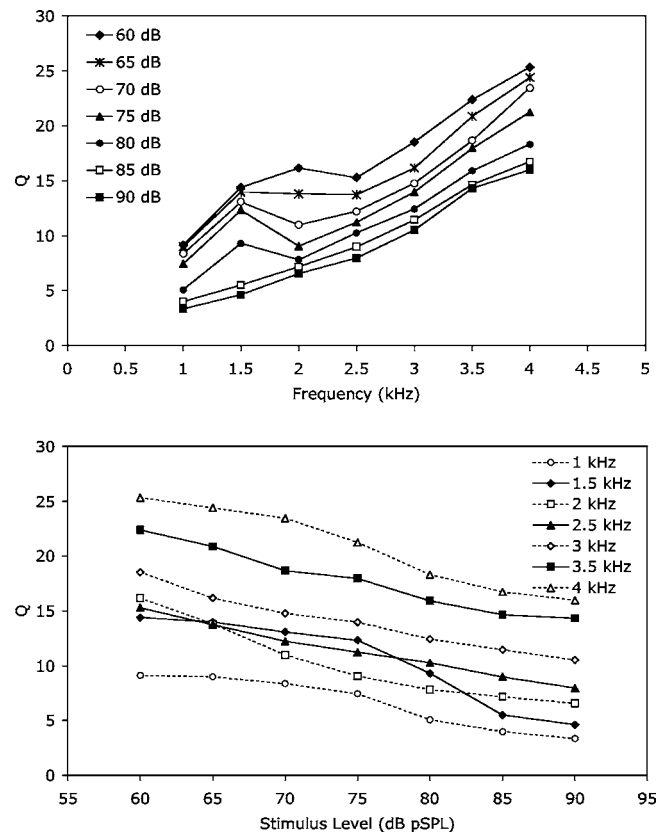
FIG. 3. Cochlear tuning estimated from the TEOAE latency data of Fig. 2, as a function of frequency (top panel) and stimulus level (bottom). A steady decrease of tuning with stimulus level is observed.

B. Tuning estimates from TEOAE latency measurements

Using the latency data of Fig. 2 in Eq. (7) one gets the cochlear quality factor estimates for different stimulus levels shown in the top panel of Fig. 3. Using Eq. (6) would give slightly different estimates of the quality factor, which are not shown here, for brevity. The regular behavior of the latency estimates of Fig. 2 corresponds to an increase of the quality factor with increasing frequency and decreasing stimulus level. A local maximum of the quality factor may be noted in Fig. 3 at low stimulus levels around the frequency of 1.5 kHz, near the maximum of the average TEOAE spectral response. In the lower panel of Fig. 3 the dependence of the quality factor on the stimulus level is shown more clearly, for each frequency band between 1 and 4 kHz.

This behavior agrees with conspicuous evidence from psychoacoustics, and other OAE-based estimates obtained studying the DPOAE suppression tuning curves (Kummer *et al.*, 1995; Gorga *et al.*, 2003) and the SFOAE phase-gradient delay (Shera *et al.*, 2002; Shera and Guinan, 2003). This increase of the quality factor with frequency had been already found analyzing neonate and adult TEOAEs at stimulus levels of $77-80$ dB (Sisto and Moleti, 2002; Moleti and Sisto, 2003; Moleti *et al.*, 2005). As mentioned earlier, a steady increase of tuning with frequency had also been observed by Shera *et al.* (2002) using both behavioral and SFOAE techniques, but the slope of the power law relation that they found was much slower (-0.3) than that of Fig. 3.

Previous psychoacoustic estimates (Glasberg and Moore, 1990) had found even slower increase of the quality factor with frequency in humans. Apart from the uncertainty associated with the model-dependent procedure used in this study to estimate cochlear tuning from latency measurements, the difference between the behavior shown in Fig. 3 and other tuning estimates could be partly influenced by the use of click stimuli.

(1) Having used an ideally flat-spectrum wide-band click stimulus, which typically decreases around 4 to 5 kHz, the latency/frequency relation is measured at approximately constant stimulus spectral density level. It is well known that psychoacoustical tuning estimates are different if measured at constant signal level or at constant masker level. In this study, Eqs. (6)–(8) explicitly refer to the quality factor of the cochlear local resonant response, as influenced by the local (within a critical band) BM excitation (in terms of rms transverse displacement). Therefore, as mentioned in Sec. III it is not clear whether, in our case, the quality factor should be measured at constant stimulus level or at constant cochlear excitation level, which cannot be done without making additional assumptions.

(2) The cochlear transmission delay implies that the click stimulus activates each cochlear place immediately after having activated the neighboring more basal places, and this asymmetry could affect the slope of the tuning/frequency relatio.

(3) Cochlear nonlinearity implies that using a click is not fully equivalent to separately using pure tone stimuli at different frequencies. In particular, intermodulation distortion components could be present, as demonstrated in guinea pigs by Yates and Withnell (1999).

For these reasons, it would be interesting to repeat these measurements using tone bursts of different bandwidth.

C. Cochlear nonlinearity and tuning dependence on stimulus level

The decrease of the TEOAE latency with increasing stimulus level corresponds, in cochlear models, to a decrease of the quality factor Q. As mentioned in Sec. III, in transmission-line cochlear models, this behavior is described by introducing a damping function, which is nonlinearly dependent on the local BM transverse displacement.

In Fig. 4 the damping coefficient derived from the TEOAE latency is shown as a function of the stimulus level. The behavior shown in Fig. 4 is in qualitative agreement with the concept of nonlinear damping. The increase of the damping coefficient with stimulus level is indeed predicted at high stimulus levels, where the quadratic nonlinear term of Eq. (8) is dominant. As previously remarked, a meaningful check of the theoretical models of the cochlear amplifier nonlinearity requires evaluating the dependence of the damping coefficient Γ on the BM excitation amplitude ξ, and not on the stimulus level. For this reason, in Fig. 5 the average damping coefficients of Fig. 4 are plotted as a function of the average TEOAE response $p(f)$ squared. Assuming that the
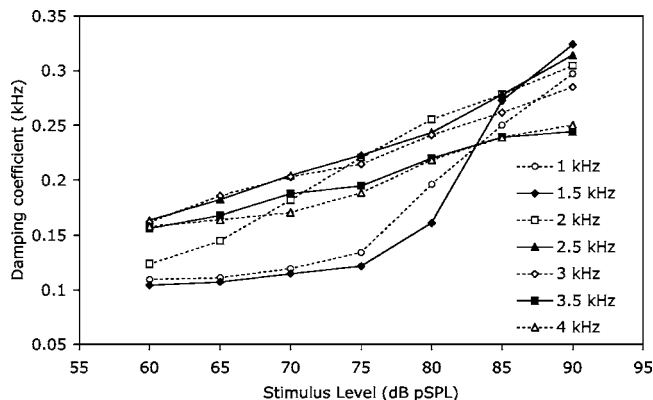
FIG. 4. Damping coefficient estimated from the TEOAE latency data of Fig. 2, as a function of the stimulus level, for each frequency band. The increase of the damping coefficient is predicted by nonlinear cochlear models, but the irregular behavior shown here across frequencies suggests that stimulus level could not be the best independent variable to describe this variation.

excitation amplitude at a given cochlear place $\xi(x(f))$ is proportional to the amplitude of the corresponding otoacoustic emission $p(f)$, a linear behavior of the relation plotted in Fig. 5 is predicted by Eq. (8). It can be noticed that the data tend indeed to cluster along well-distinct straight lines, which seems to confirm the overall coherence of the proposed interpretative framework.

The positive value of the intercept, which would be associated with the coefficient B of Eq. (8), favors a nonlinear oscillator model of the type proposed by Sisto and Moleti (1999), while a Van der Pol oscillator model would predict a negative intercept. As the intercept of the straight line corresponds in the model to the damping coefficient in the linear low-stimulus-level regime, it is also interesting to note that these values of B are in the same range as those obtained by Sisto et al. (2001), measuring the time constant of the exponential decay of long-lasting OAEs after excitation by click stimuli.
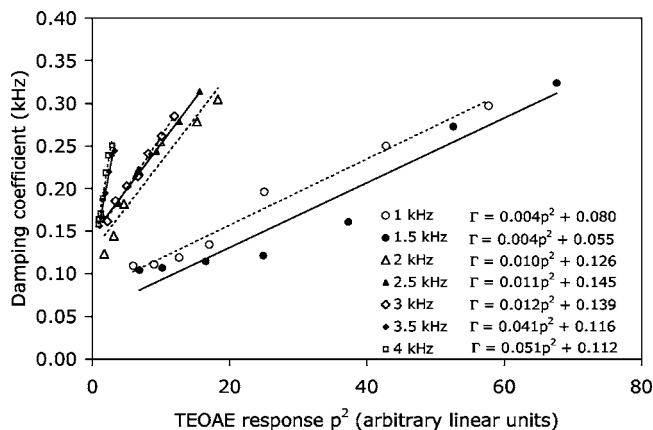
FIG. 5. Average damping coefficient of each frequency band plotted vs the square of the average response amplitude in the corresponding band. The damping coefficient is expressed in kilohertz, as the frequency, to help comparison. The straight line fit functions are also shown for each frequency. Note how the data cluster separately along straight lines for each frequency, differently from Fig. 4, and that the intercepts of the fit functions are always positive, as predicted by Eq. (8).

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

R. Sisto and A. Moleti: Otoacoustic emission latency and tuning   2189

V. CONCLUSIONS

The TEOAE latency can be effectively measured as a function of frequency using a time-frequency technique based on the wavelet transform. TEOAE latency data can be used to get objective, even if model-dependent, estimates of cochlear tuning. The tuning estimates obtained at different stimulus levels show that the quality factor of the cochlear filters (1) increases with frequency and (2) decreases with the stimulus level, confirming previous results based on OAE data analysis and psychoacoustical studies. The observed dependence on the stimulus level is compatible with the notion of nonlinear damping increasing quadratically with the amplitude of the cochlear local transverse displacement, including also a linear positive damping term, in agreement with the predictions of transmission-line cochlear models. This work confirms that linking different findings from OAE experiments with cochlear models can help in getting a coherent schematization of the inner ear mechanisms.

Furst, M., and Lapid, M. (**1988**). "A cochlear model for acoustic emissions," J. Acoust. Soc. Am. **84**, 222–229.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Gorga, M. P., Neely, S. T., Dierking, D. M., Dorn, P. A., Hoover, B. M., and Fitzpatrick, D. F. (**2003**). "Distortion product otoacoustic emission suppression tuning curves in normal-hearing and hearing-impaired human ears," J. Acoust. Soc. Am. **114**, 263–278.

Greenwood, D. D. (**1990**). "A cochlear frequency position function for several species −29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Jedrzejczak, W. W., Blinowska, K. J., Konopka, W., Grzanka, A., and Durka, P. J. (**2004**). "Identification of otoacoustic emission components by means of adaptive approximations," J. Acoust. Soc. Am. **115**, 2148–2158.

Kummer, P., Janssen, T., and Arnold, W. (**1995**). "Suppression tuning characteristics of the 2f1-f2 distortion product otoacoustic emission in humans," J. Acoust. Soc. Am. **98**, 197–210.

Mallat, S. G., and Zhang, Z. (**1993**). "Matching pursuit with time-frequency dictionaries," IEEE Trans. Signal Process. **41**, 3397–3415.

Martin, G. K., Jassir, D., Stagner, B. B., Whitehead, M. L., and Lonsbury-Martin, B. L. (**1998**). "Locus of generation for the 2f1-f2 vs 2f2-f1 distortion-product otoacoustic emissions in normal-hearing humans revealed by suppression tuning, onset latencies, and amplitude correlations," J. Acoust. Soc. Am. **103**, 1957–1971.

Moleti, A., and Sisto, R. (**2003**). "Objective estimates of cochlear tuning by otoacoustic emission analysis," J. Acoust. Soc. Am. **113**, 423–429.

Moleti, A., Sisto, R., Tognola, G., Parazzini, M., Ravazzani, P., and Grandori, F. (**2005**). "Otoacoustic emission latency, cochlear tuning, and hearing functionality in neonates," J. Acoust. Soc. Am. **118**, 1576–1584.

Moulin, A., and Kemp, D. T. (**1996**). "Multicomponent acoustic distortion product otoacoustic emission phase in humans. II. Implications for distortion product otoacoustic emissions generation," J. Acoust. Soc. Am. **100**, 1640–1662.

Neely, S. T., Norton, S. J., Gorga, M. P., and Jesteadt, W. (**1988**). "Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli," J. Acoust. Soc. Am. **83**, 652–656.

Norton, S. J., and Neely, S. T. (**1987**). "Tone-burst-evoked otoacoustic emissions from normal-hearing subjects," J. Acoust. Soc. Am. **81**, 1860–1872.

Prieve, B. A., Gorga, M. P., and Neely, S. T. (**1996**). "Click- and tone-burst-evoked otoacoustic emissions in normal-hearing and hearing-impaired ears," J. Acoust. Soc. Am. **99**, 3077–3086.

Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2067.

Schoonhoven, R., Prijs, V. F., and Schneider, S. (**2001**). "DPOAE group delays versus electrophysiological measures of cochlear delay in normal human ears," J. Acoust. Soc. Am. **109**, 1503–1512.

Shera, C. A., and Guinan, J. J., Jr. (**2003**). "Stimulus-frequency emission group delay: A test of coherent reflection filtering and a window on cochlear tuning," J. Acoust. Soc. Am. **113**, 2762–2772.

Shera, C. A., Guinan, J. J., Jr., and Oxenham, A. J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.

Shera, C. A., Tubis, A., and Talmadge, C. L. (**2005**). "Coherent reflection in a two-dimensional cochlea: Short-wave versus long-wave scattering in the generation of reflection-source otoacoustic emissions," J. Acoust. Soc. Am. **118**, 287–313.

Sisto, R., and Moleti, A. (**1999**). "Modeling otoacoustic emissions by active non linear oscillators," J. Acoust. Soc. Am. **106**, 1893–1906.

Sisto, R., and Moleti, A. (**2002**). "On the frequency dependence of the otoacoustic emission latency in hypoacoustic and normal ears," J. Acoust. Soc. Am. **111**, 297–308.

Sisto, R., Moleti, A., and Lucertini, M. (**2001**). "Spontaneous otoacoustic emissions and relaxation dynamics of long decay time OAEs in audiometrically normal and impaired subjects," J. Acoust. Soc. Am. **109**, 638–647.

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.

Tognola, G., Grandori, F., and Ravazzani, P. (**1997**). "Time-frequency distributions of click-evoked otoacoustic emissions," Hear. Res. **106**, 112–122.

van Hengel, P. W. J., Duifhuis, H., and van den Raadt, M. P. M. G. (**1996**). "Spatial periodicity in the cochlea: The result of interaction of spontaneous emissions?," J. Acoust. Soc. Am. **99**, 3566–3571.

Wable, J., Collet, L., and Chery-Croze, S. (**1996**). "Phase delay measurements of distortion product otoacoustic emissions at 2f1-f2 and 2f2-f1 in human ears," J. Acoust. Soc. Am. **100**, 2228–2235.

Whitehead, M. L., Stagner, B. B., Martin, G. K., and Lonsbury-Martin, B. L. (**1996**). "Visualization of the onset of distortion-product otoacoustic emissions, and measurement of their latency," J. Acoust. Soc. Am. **100**, 1663–1679.

Wit, H. P., van Dijk, P., and Avan, P. (**1994**). "Wavelet analysis of real ear and synthesized click evoked otoacoustic emissions," Hear. Res. **73**, 141–147.

Yates, G. R., and Withnell, R. H. (**1999**). "The role of intermodulation distortion in transient-evoked otoacoustic emissions," Hear. Res. **136**, 49–64.

Zweig, G. (**1991**). "Finding the impedance of the organ of Corti," J. Acoust. Soc. Am. **89**, 1229–1254.

Zweig, G., and Shera, C. A. (**1995**). "The origin of periodicity in the spectrum of otoacoustic emissions," J. Acoust. Soc. Am. **98**, 2018–2047.

A comparative study of distortion-product-otoacoustic-emission fine structure in human newborns and adults with normal hearing

Sumitrajit Dhar
Hugh Knowles Center, Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, Illinois 60208

Carolina Abdala
Children's Auditory Research and Evaluation Center, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

Distortion product otoacoustic emissions (DPOAE) measured in human newborns are not adult-like. More than a decade of work from various investigators has created a well-developed body of evidence describing these differences but the putative anatomy or physiology has only been partially explained. Recently, Abdala and Keefe [J. Acoust. Soc. Am. **120**, 3832–3842 (2006)] have identified outer and middle ear immaturities that at least partially describe the differences observed between newborn and adult input–output functions and suppression tuning curves. DPOAE fine structure characteristics and their maturation have not been examined to any extent in the literature. Fine structure characteristics in two groups of ten newborns and young adults with normal hearing sensitivity are compared here. Consistent with previous reports, the newborns show higher DPOAE levels; greater fine structure depth and wider fine structure spacing is also observed in the newborns. Differences in fine structure morphology are also observed between the two age groups. While some of these findings are attributable to an immature outer and middle ear system in the newborns, it is argued that some observed differences in fine structure characteristics might be due to remnant immaturities in passive motion of the basilar membrane in the newborn cochlea.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2770544]

I. INTRODUCTION

Otoacoustic emissions (OAEs) are sounds generated in the inner ear that can be recorded in the ear canal (Kemp, 1978). When evoked using two simultaneous pure tones, distortion product (DP) OAEs can be recorded at frequencies mathematically related to the stimulus frequencies (Kemp, 1979a). The DPOAE at the frequency $2f_1 - f_2$ (f_1 and f_2 represent the frequencies of the stimulus tones, $f_2 > f_1$) is the most commonly studied and used clinically for detection of hearing loss (e.g., Dorn *et al.*, 1999). Although DPOAEs are widely used as a noninvasive assay in studying various aspects of peripheral auditory physiology and biophysics, our particular interest here is in their use as a tool to investigate maturation of the auditory periphery in humans (Abdala, 1998, 2001a, b, 2003; Abdala and Chatterjee, 2003; Abdala, 2004; Abdala and Keefe, 2006; Keefe and Abdala, 2007) as well as laboratory animals (Mills, 2004).

A. DPOAEs to study maturation of auditory peripheral function

A decade of work from various laboratories has found that DPOAE-based measures of cochlear function are not completely adult-like in human infants (Brown *et al.*, 1994; Abdala, 1998; Lasky, 1998; Abdala, 2001b, 2004). Brown and colleagues (1994) reported immature cochlear filtering in newborns at 4000 Hz, as measured by DPOAE f_2/f_1-ratio functions. DPOAE input/output (I/O) functions are not adult-like in infants either, and show immature saturation characteristics (Lasky, 1998; Abdala, 2000). DPOAE ipsilateral suppression tuning curves (STCs) at $f_2 = 6$ kHz have been found to be non-adult-like in very premature infants (30 weeks, postconceptional age), term-born infants and older infants through 6 months of age (Abdala *et al.*, 1996; Abdala, 1998, 2004; Abdala *et al.*, 2007). Typically, infant DPOAE STCs are significantly narrower in width, steeper on the low-frequency flank and have a sharper tip region than adult STCs. Additionally, the growth of DPOAE suppression (for low-frequency suppressor tones only) is shallower for infants than adults.

Clearly, DPOAE-based measures of peripheral auditory function recorded from infants are not adult-like in the high-frequency range (4–6 kHz) and they remain immature well into the post-natal period. The source(s) of these immaturities has not been fully specified, although recent evidence suggests that immaturities in the outer and middle ear system contribute significantly to the non-adult-like DPOAE features observed in human infants (Abdala and Keefe, 2006; Keefe and Abdala, 2007). It is not clear whether cochlear or medial olivocochlear sources contribute to the residual immaturities not attributable to the outer and middle ear system. Relatively recent confirmation that the DPOAE is comprised of multiple components arising from different

locations in the cochlea, provides yet another paradigm to study maturation of auditory peripheral function in humans.

B. DPOAE fine structure and components

The observation of a pseudo-periodic pattern of alternating maxima and minima in high-resolution recordings of DPOAEs, now known as fine structure, led to the initial prediction that the ear-canal DPOAE was comprised of more than one component (Kim, 1980). Following extensive theoretical and experimental work, the "two-source" model of DPOAEs is now well accepted for DPs where the characteristic frequency (CF) region is apical to that of the stimulus tones. In brief, for DPOAEs with CF regions on the basilar membrane apical to those of stimulus tones (i.e., $f_2 > f_1 > f_{dp}$), the initial DPOAE energy is generated in the overlap region between the traveling wave patterns evoked by the stimulus tones. This initial distortion energy is distributed bidirectionally with a portion propagating outwards to the ear canal and another portion propagating inwards (toward the apex of the cochlea). This second, inward propagating portion of the distortion energy reaches its CF region, is arguably affected by active physiological elements of the cochlea locally, and is returned to the ear canal as a second DPOAE component. The "interference" created by this second component and its impact (constructive or destructive) on DPOAE level can be observed in DPOAE fine structure patterns. The reader is directed to other sources for analytic treatments of this model of DPOAE generation (Talmadge *et al.*, 1998; Mauermann *et al.*, 1999a, b; Talmadge *et al.*, 1999).

Confirmation of two DPOAE components from two disparate locations in the cochlea comes from a variety of experiments. A suppressor of appropriate magnitude placed in close proximity to the $2f_1 - f_2$ CF region eliminates/alters the component from this region, thereby reducing its contribution to the ear canal DPOAE and eliminating fine structure (Heitmann *et al.*, 1998; Talmadge *et al.*, 1999; Dhar and Shaffer, 2004). Fine structure is also eliminated when the CF region but not the overlap region falls in an audiometric notch, or region of hearing loss (Mauermann *et al.*, 1999b); the overlap component remains unaffected but the interference is eliminated due to the absence of a second component from the DP CF region. Finally, interference between the two components is also evident in time domain recordings when one stimulus tone is pulsed to alternately turn DPOAE generation on and off (Talmadge *et al.*, 1999). As the stimulus is turned on, the overlap component is recorded in isolation in the ear canal for a brief duration, whereas the second component travels to the DP CF region and is returned from there with a finite time delay compared to the component from the overlap region. By choosing the appropriate phase relationship between the two components, a cancellation notch can be observed in the ear canal at either signal onset or offset.

The DPOAE components from the overlap and DP CF regions have also been classified as *wave* and *place* fixed emissions (Knight and Kemp, 2000, 2001), as well as *nonlinear* and *reflection* components (Shera and Guinan, 1999; Shera and Zweig, 1991). Each of these models proposes a different mechanism of OAE generation and backward propagation. The mechanism notwithstanding, the presence of two (or more) DPOAE components in the ear canal signal is well established and accepted today.

The individual characteristics of these DPOAE components as a function of stimulus frequency ratio (Knight and Kemp, 2001; Dhar *et al.*, 2005) and stimulus level (Konrad-Martin *et al.*, 2001) have been of recent interest. The differential vulnerability of these two components has also been examined to a limited extent (Engdahl and Kemp, 1996). However the *developmental* aspects of DPOAE components and fine structure characteristics have remained largely unexplored. Here we report on a comparison of DPOAE components and their interaction, between human newborn and adult ears, as assessed by the fine structure of the ear canal DPOAE signal. To the best of our knowledge, such a comparison is not available in the peer-reviewed literature as of this writing.

II. METHODS

A comparison of DPOAE fine structure characteristics across two sets of independently recorded data is presented here. DPOAEs in ten newborns were recorded at the House Ear Institute in Los Angeles, California, whereas the results from ten young adults between the ages of 18 and 24 years were recorded at Northwestern University in Evanston, Illinois. The two data sets were recorded in accordance with the institutional review board regulations at the respective institutions and each data set was thoroughly de-identified before sharing with outside research personnel. In the following we provide separate descriptions of hardware, software for data recording, and initial estimation of DPOAE level and phase for each data set. The common analyses used to extract DPOAE fine structure parameters from both data sets are described as well.

A. Newborn data set

Ten term-born, normal-hearing human infants (mean gestational age at birth=38.6 weeks) served as subjects in this study, following acquisition of informed parental consent. They were tested within 72 h after birth. Nine left and one right ear were tested from five male and five female newborns with an average birth weight of 3200 g. Infants were wheeled in open isolettes from the mother's postpartum room to the Infant Auditory Research Laboratory in Women's and Children's Hospital, University of Southern California, Los Angeles County (USC+LAC). They were fed and changed if necessary prior to testing and once asleep, the DPOAE probe was placed at the entrance of the ear canal for testing.

A custom-designed DPOAE acquisition system (SuprDP) was used to generate stimuli and acquire data under the control of custom software using a 48 000 Hz sampling rate. The data acquisition hardware was based on an audio processor module developed by The House Ear Institute Engineering Department. The hardware includes two-channel D/A, two-channel A/D and a digital-signal processor (all 24-bit) as well as an analog high-pass filter (12 dB/oct.; 700 Hz

high-pass cutoff; note that this affected three data points where the DPOAE frequency was below 700 Hz). The data acquisition system was connected to an Etymotic Research ER-10C probe microphone. The ER-10C contains two output transducers and a low-noise microphone. The two stimulus tones were generated by the DSP processor and presented to the infants' ear canals using separate transducers. The microphone signal was high-pass filtered before being sampled by the A/D converter.

Twenty sweeps of the microphone signal were averaged by the DSP processor and comprised one block of data. Sweeps were accepted into a block only when the estimated root-mean-squared level in that sweep did not exceed a user-controlled artifact rejection threshold. Additionally, a block of data could be rejected by the user if the mean noise floor was >0 dB. A data point was eliminated automatically if a spike in noise occurred (defined ≥15 dB increase in noise from one point to an adjacent point) and SNR was <10 dB. A minimum of 6 and a maximum of 12 acceptable blocks of data were averaged to compute the final DPOAE amplitude or the DPOAE grand average. DPOAE level and phase were extracted from recordings using fixed stimulus levels of 65–55 dB SPL ($L_1 > L_2$) and f_2/f_1 ratio of 1.2. These stimulus parameters were used for both the newborn and adult groups as they are typically used in the clinic and may not be optimized for measuring the most robust fine structure in either group. For a first attempt at examining fine structure in human infants, moderate-low level of stimuli were used (rather than very low levels that might have enhanced fine structure), in order to produce robust DPOAE levels and sufficient signal-to-noise ratio (SNR) to observe an unequivocal response. DPOAEs were recorded between f_2 frequencies of 996–4020 Hz in 11.7 Hz intervals.

Intermodulation distortion produced by the recording system at $2f_1 - f_2$ was on average, −25 dB SPL. The recording system mean noise floor ranged between −23 and −30 dB SPL. An *in situ* calibration procedure was conducted on both output transducers before each subject was tested. A chirp tone (swept-frequency signal from 10 to 10 000 Hz) with fixed voltage was presented to the transducer and the resulting SPL recorded in the ear canal. Based on this information, an equalization of output levels was performed for each subject to achieve target stimulus levels across test frequencies.

B. Adult data set

Individuals responding to advertisement fliers on the campus of Northwestern University in Evanston, Illinois were screened for inclusion in the experiment. Individuals, with hearing thresholds better than 20 dB HL (ANSI, 1996) between 0.25 and 8 kHz, negative history of otologic disease and noise exposure, and normal middle ear function indicated by a type A tympanogram, defined by static compliance between 0.4 and 1.5 cm³ and peak pressure between ±150 daPa were included for detailed DPOAE recordings.

Signal generation and recording was controlled using custom software on an Apple Macintosh G4 computer via a MOTU 828 Mk II firewire I/O device (24 bit/44 100 Hz).

The stimulus tones were passed through a Behringer 8000 ProXL headphone amplifier and presented to the subjects' ear canals via MB Quartz HB13.x transducers. The output of the transducers was coupled to the ear canal through the probe assembly of an Etymotic Research ER10B microphone. The output of the microphone and its preamplifier was passed through a Stanford Research SR560 low-noise voltage amplifier with band pass filtering between 0.3 and 20 kHz. The amplified output from the SR560 was digitized by the MOTU and stored on disk.

DPOAE recordings were made between f_2 frequencies of ~782 Hz ($2f_1 - f_2 = 500$ Hz) and ~18 700 Hz ($2f_1 - f_2 = 12 000$ Hz) using stimulus levels of 65 (L_1) and 55 (L_2) dB SPL and a stimulus frequency ratio of (f_2/f_1) 1.22. In the analyses presented here, DPOAE data for f_2 frequencies between 1000 and 4000 Hz are presented to maintain consistency with the newborn group. In this frequency range, the stimulus tones were swept at a rate of 8 s/octave. Four to six such sweeps were averaged and DPOAE level, phase, and noise floor estimates were made using a least-square-fit algorithm (Long *et al.*, 2004). This analysis yielded DPOAE level, noise floor, and phase estimates at every 2 Hz around $f_2 = 1000$ Hz and every 6 Hz around $f_2 = 4000$ Hz. The stimulus levels were calibrated and system distortion measured in a B&K 4157 ear simulator using a B&K 4134 microphone. System distortion was approximately −35 dB SPL for the stimulus levels used in these recordings.

C. Characterization of fine structure

Data from both age groups were processed identically to estimate fine structure parameters of depth and spacing. Although the data from the two age groups were recorded using different frequency resolutions, even the lower resolution (11.7 Hz used in the infants) would have been sufficient to accurately characterize the DPOAE fine structure spacing reported from adult ears in the literature. The median value for every three successive data points was computed for DPOAE level and noise floor. Data points where the SNR between the level and noise-floor medians did not meet the criterion value of 6 dB were eliminated. Comparing the SNR of medians rather than single data points allowed the retention of fine structure minima where the local SNR was unfavorable.

Fine structure maxima and minima were identified based on the first and second derivatives of the DPOAE level function and the relationship between them. Data points where the first derivative was equal to zero were identified as extrema (maximum or minimum), and then further classified as a maximum or minimum based on the second derivative being negative or positive, respectively. Data points where the second derivative was equal to zero were marked as inflection points (Dhar *et al.*, 2002). Fine-structure depth for each period was computed as: $FS_{depth} = 20 \log_{10}(P_{max}/P_{avg(min)})$, where P_{max} is the DPOAE amplitude at a maximum and $P_{avg(min)}$ is the average DPOAE amplitude of the preceding and following minima. The frequency spacing of fine structure was computed as the ratio $f/\Delta f$, where f is the geometric mean between two adjacent minima and Δf is the frequency separation between them (Shera, 2003). The analysis
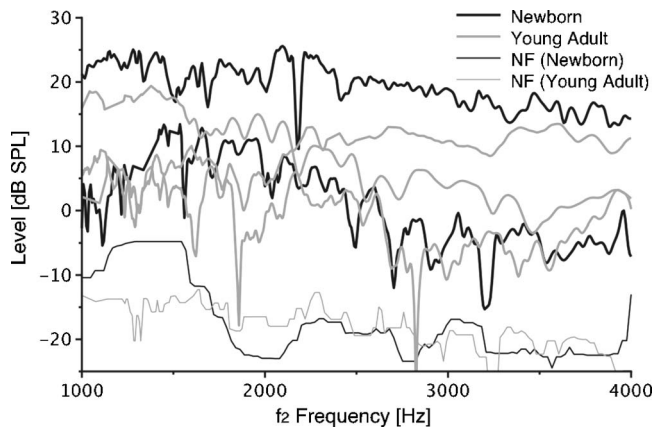
FIG. 1. DPOAE level from *two* newborn and *three* adult subjects in black and gray lines, respectively. Averaged noise floors for each group are presented using thinner lines in the same color. These subjects from each group were selected to represent the general limits of DPOAE level in each group.

FIG. 2. Comparison of DPOAE fine structure in one member from each age group. These two subjects were chosen to represent examples of "deep" fine structure from each group. The frequency range marked by the horizontal dashed line allows direct comparison of approximately three fine structure periods in the two subjects. The noise floor from each subject is also displayed.

described previously was duplicated for DPOAE group delay (negative of the slope of phase). Only fine structure periods where depth was greater than 3 dB were included in further analyses, unless matching periods could be observed in the level and group delay data. Only 3 out of a total of 162 qualified fine structure periods were less than 3 dB deep. Further, to be included in the analyses, an odd number of inflection points had to be observed between a maximum and a minimum. Each fine structure period was classified as "log-sine" or "non-log-sine" based on the number of inflection points observed in each half period (see Dhar *et al.*, 2002, for details).

DPOAE level as well as fine structure prevalence, depth and spacing were compared across age groups, frequency, and gender in independent multiway analyses of variance (ANOVAs) using the statistical computing package R (R-Development-Core-Team, 2006). For analysis and display related to overall DPOAE level, f_2 frequency was used. Frequency was transposed to $2f_1 - f_2$ for all fine structure related analyses and display. All dependent variables were assumed to be normally distributed and averaged across a range of frequencies to yield four values for nominal test frequencies—then treated as a repeated measure in the analyses. The frequency range over which each dependant variable was averaged is specified in the following section.

III. RESULTS

DPOAE levels from two newborn and three young adult subjects along with averaged noise floors are displayed in Fig. 1. The black lines represent data from the newborns while the gray lines represent data from the young adults. These subjects were selected from each age group to demonstrate the range of DPOAE levels observed. DPOAE levels recorded from one of the newborns are considerably higher across the entire frequency range, even if data from two young adults are combined for comparison in low (<2 kHz) and high (>2 kHz) frequency ranges. The lower limit of the range of DPOAE levels appears to be equivalent between the two age groups.

DPOAE levels as a function of frequency from one subject in each age group are displayed in Fig. 2 along with the corresponding noise floors. These subjects were chosen based on the presence of pronounced fine structure. Distinct fine structure periods are observed across the entire $(2f_1-f_2)$ frequency range for the newborn. In the young adult, such distinct fine structure is observed only between the frequencies of 1000 and 1500 Hz (marked in Fig. 2 for easy comparison). In this frequency range, the fine structure appears to be deeper in the newborn. Note that the general DPOAE level is equivalent in these two subjects across the entire frequency range.

Figure 3 allows comparison of mean DPOAE levels across age groups and sex. Three different estimates of DPOAE level are displayed for the newborns along with the noise floor in panel A. The solid circles represent the DPOAE level measured at the exact f_2 frequency in ten newborns. The gray circles represent DPOAE levels averaged over a span of 1000 Hz centered at each f_2 frequency in each newborn and then averaged across subjects. The frequency limits of recorded data in effect cause the averages to be over 500 Hz for the f_2 frequencies of 1000 and 4000 Hz. The open circles represent averages computed over a distance of 2 mm on the basilar membrane (Greenwood, 1990) centered around the nominal test frequencies in each subject and then averaged across subjects. The solid squares represent the average noise floor computed over a span of 2 mm also and the error bars represent ±1 standard deviation. The mean DPOAE levels following the 1000 Hz and 2 mm averaging are dissimilar at 1000 Hz but increasingly similar with increasing frequency. The same frequency range spans a greater distance on the basilar membrane with decreasing frequency. Thus, averages taken over a fixed frequency range in different regions of the cochlea are not equivalent. We choose to use the DPOAE levels averaged over a fixed distance on the basilar membrane (2 mm) for all further comparisons. In doing so we assume that Greenwood's map is valid in the infants, i.e., the length of the infant cochlea is
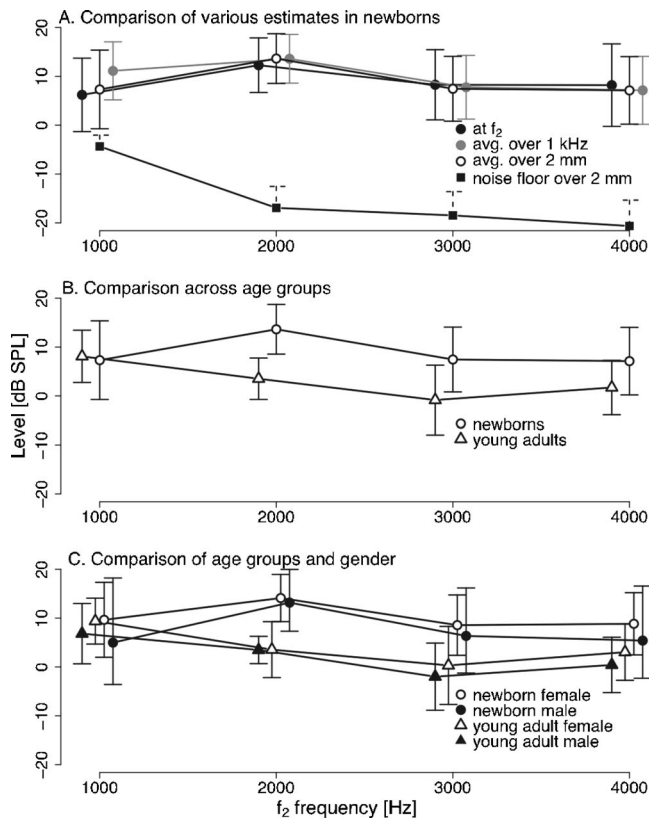
FIG. 3. Comparisons of various estimates of DPOAE levels as a function of f_2 frequency. Error bars represent ±1 standard deviation. (A) Comparison of average DPOAE level estimated using three different methods in ten newborns along with an estimate of the average noise floor (see the text). (B) Average DPOAE levels for ten newborns and ten young adults as a function of frequency. (C) Average DPOAE levels for each age group separated by sex. Symbols are "jittered" around the nominal frequency points on the abscissa to enhance visual clarity.

adult like. Using averages across fixed distances equalizes, to some extent, the number of data points representing different portions of the cochlea.

In panel B of Fig. 3, averaged DPOAE levels are compared across age group. Mean DPOAE levels are higher in the newborns at all frequencies except 1000 Hz. Overall, the grand average of DPOAE level was higher in the newborns by approximately 7 dB. The main effects of age group and frequency on DPOAE level were found to be statistically significant (see Table I). The data are further divided by sex in panel C of Fig. 3. In both age groups, there is a trend toward higher DPOAE levels in females even though no

FIG. 4. Total number of fine structure periods observed in all (left) ten newborns and (right) ten young adults. Data from female and male subjects are presented in open and shaded bars, respectively. The main effect of sex was statistically significant (see Table I).

main effect of sex was observed in the analysis of variance. This sex difference is the smallest at 2 kHz in both age groups.

A count of qualified fine structure periods observed in male and female subjects belonging to the two age groups is displayed in Fig. 4. Female newborns had the largest number of fine structure periods followed by newborn males, young adult females, and young adult males. The main effect for age group was statistically significant as was the interaction between sex and age group (see Table I). The difference in the total number of fine structure periods observed was greater between male and female ears in the newborns.

Fine structure depth was significantly greater in newborns than young adults (Fig. 5 and Table I). Fewer fine structure periods were observed at $2f_1 - f_2$ frequencies below 1000 Hz, primarily due to higher noise floors at these frequencies. Female ears showed significantly deeper fine structure in both age groups (Table I and Fig. 6). Figure 6 shows the distribution of fine structure depth in both age groups separated by sex. In newborn ears, the distributions in female and male ears show peaks between 4 and 8 dB, with a gradual decline beyond the peak. In contrast, the peak of the distribution in adults is between 2 and 6 dB and the decline in the distributions is more drastic beyond the maximum for young adults. Notably, no fine structure periods were observed in one young adult ear.

TABLE I. Results of multiway ANOVA with age group, sex, and frequency as the independent variables and either DPOAE level, fine structure prevalence, depth, or fine structure spacing as the dependent variables. Interactions between age group and frequency as well as between sex and frequency were not statistically significant for any dependent variable. To arrive at the mean DPOAE level estimates, a minimum of 20 data points (around 1 kHz) and a maximum of 85 data points (around 4 kHz) were averaged. The number of data points used in the analyses involving fine structure parameters at any frequency depended on prevalence and are shown in Figs. 4–8.

Variables	DPOAE Level	Fine structure prevalence	Fine structure depth	Fine structure spacing
Age Group	$p < 0.01$, $F = 15.34$	$p < 0.01$, $F = 26.62$	$p < 0.01$, $F = 31.67$	$p < 0.01$, $F = 34.77$
Sex			$p < 0.05$, $F = 5.4$	
Frequency	$p < 0.05$, $F = 5.29$			$p < 0.01$, $F = 70.26$
Age Group×Sex		$p < 0.01$, $F = 61.34$		
Age Group×Sex×Frequency				$p < 0.01$, $F = 16.64$

FIG. 5. Fine structure depth as a function of $2f_1-f_2$ frequency. Open circles and gray triangles represent individual data points from newborn and young adult subjects, respectively. Mean values of fine structure computed over 500 Hz ranges are displayed using the black circles and triangles for the newborns and young adults, respectively. The symbols representing the mean values are jittered along the abscissa for visual clarity. The error bars represent ±1 standard deviation. The main effect of age group was statistically significant (see Table I for details).

Spacing, computed as $f/\Delta f$, in newborns and young adults along with averages computed over 500 Hz ranges are presented in Fig. 7. A higher value of $f/\Delta f$ at any given frequency represents narrower fine structure spacing. Overall, fine structure was significantly narrower in young adults than newborns with a significant effect of frequency (Table I). A significant interaction between age, sex, and frequency was also observed. Distribution of fine structure spacing in the two age groups separated by sex is presented in Fig. 8. Note the clustering around $f/\Delta f=16$ in the young adults. Spacing in the newborn ears is more distributed at $f/\Delta f$ values less than 16.

The occurrence of log-sine and non-log-sine fine structure morphology in the two age groups separated by sex is

displayed in Fig. 9. Examples of different types of patterns are presented in the top half of Fig. 9. White portions of the bars in the bottom half of the figure represent the number of log-sine fine structure periods observed, with the gray portions representing the number of non-log-sine periods. Fine structure periods were overwhelmingly non-log-sine in the newborns with 1 period of log-sine fine structure observed in male and female newborn ears each. In contrast, the majority of fine structure periods in the young adults were of the log-sine variety.

IV. DISCUSSION

Briefly, our results demonstrate significantly greater DPOAE levels ($p<0.01$) in newborns compared to young
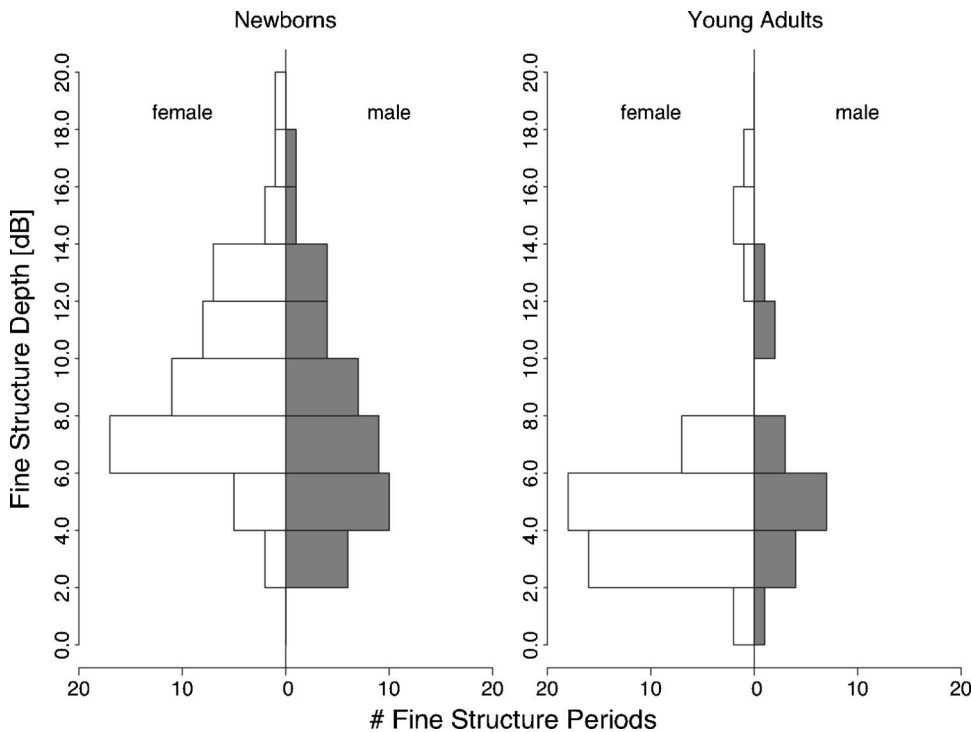
FIG. 6. Distribution of fine structure depth for the two age groups separated by sex. (Left) Data from the newborns with female and male subjects represented using open and gray bars, respectively. (Right) The format is repeated for the young adult subjects. Fine structure periods are grouped in 2 dB bins based on the measured depth. The main effect of sex was statistically significant (see Table I for details).
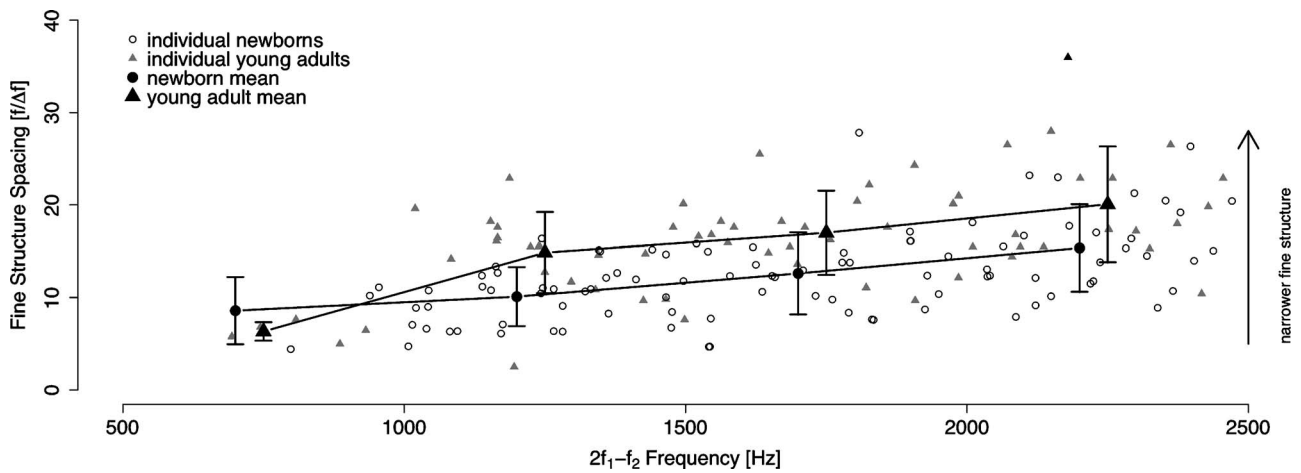
FIG. 7. Fine structure spacing computed as $f/\Delta f$ (see the text for explanation) as a function of $2f_1-f_2$ frequency. The format is identical to that of Fig. 5. Larger numbers along the ordinate represent narrower fine structure periods. The main effect of age group was statistically significant (see Table I for details).

adults, in agreement with previous results. We also report fine structure to be more prevalent, deeper, and spacing to be wider in newborns ($p < 0.01$ in each case). Fine structure depth was greater in females of either age group ($p < 0.01$). The distribution of $f/\Delta f$ was not clustered around any specific value and fine structure morphology was overwhelmingly non-log-sine in the newborns.

A. Cochlear maturation

Developmental timelines for the middle and inner ears of humans have, of necessity, commonly been obtained through indirect measures (e.g., Don *et al.*, 1993; Abdala, 1996; Keefe and Abdala, 2007). In contrast, more direct

measures can be applied to laboratory animals in the study of developmental patterns and an advantage can be gained by informed application of these findings to humans. The gerbil is one of the most extensively studied species with regards to both anatomical and physiological development of the peripheral auditory system (Harris and Dallos, 1984; McGuirt *et al.*, 1995; Mills and Rubel, 1996; Overstreet and Ruggero, 2002; Overstreet *et al.*, 2002; Mills, 2004).

The developmental status of cochlear potentials (Harris and Dallos, 1984; Yancey and Dallos, 1985; McGuirt *et al.*, 1995), basilar membrane motion (Overstreet *et al.*, 2002), cochlear amplifier gain as measured by DPOAEs (Mills and Rubel, 1994, 1996), the neural response measured at the spi-
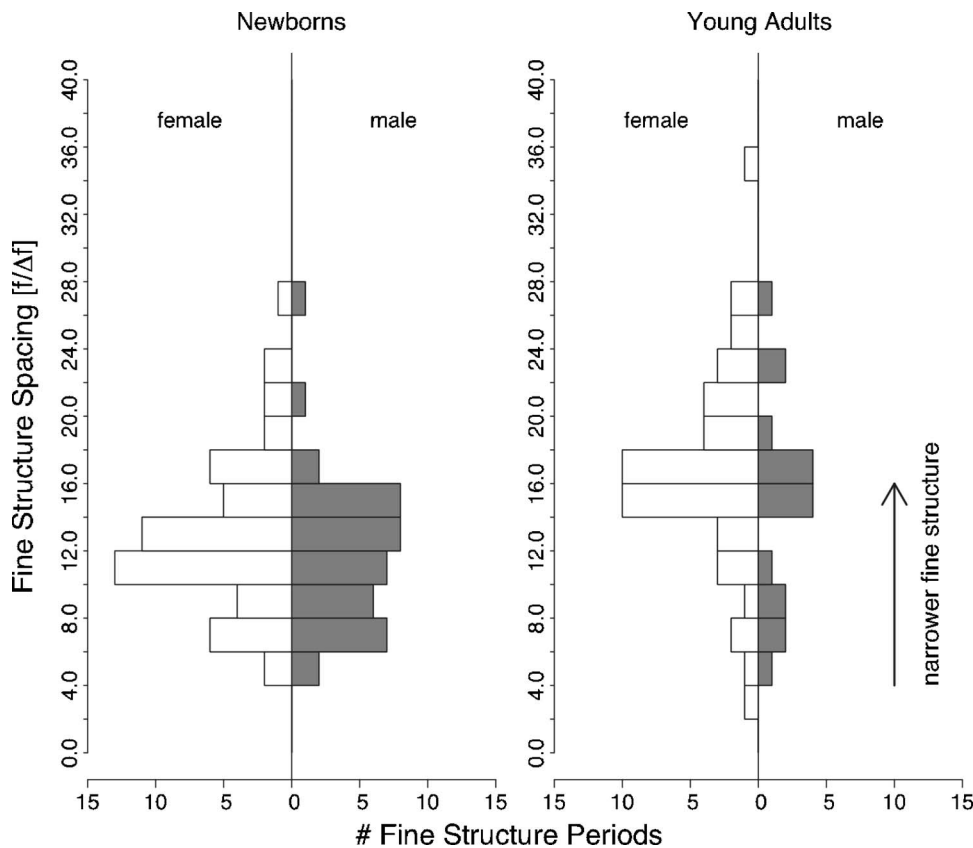
FIG. 8. Distribution of fine structure spacing in each age group and sex. The format is identical to that of Fig. 6.
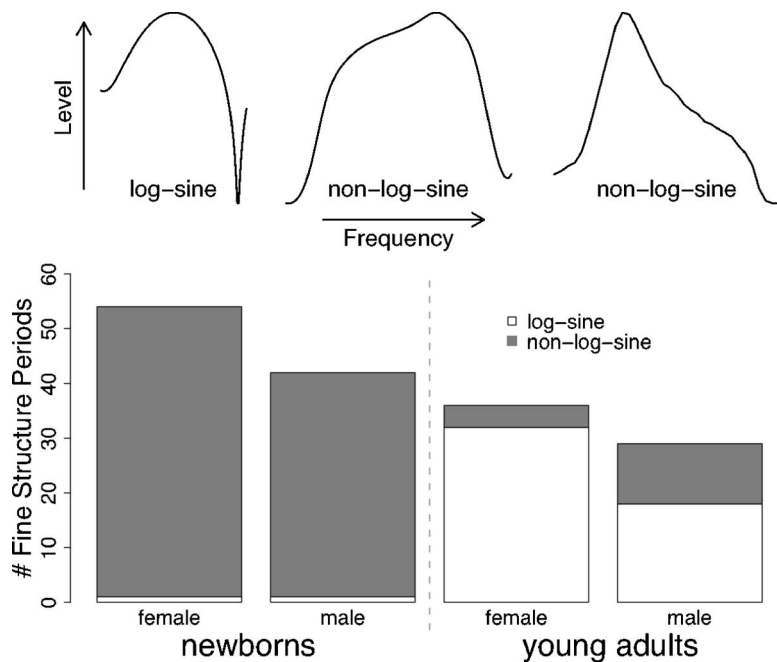
FIG. 9. Prevalence of log-sine and non-log-sine fine structure periods in each age group and sex. (Top) Examples of each type of fine structure period. (Bottom) The bars represent the total number of fine structure periods observed in each age group divided by sex. The white and gray portions within each bar represents the number of log-sine and non-log-sine periods observed in each group, respectively.

ral ganglion (Echteler *et al.*, 1989) and the auditory evoked potentials of the brain stem (Smith and Kraus, 1987) have been examined in the gerbil. The distinct advantage of many of these observations is that the developmental aspects of the cochlea can be studied independently of the concurrently maturing outer and middle ear system provided the appropriate experimental considerations are made (Harris and Dallos, 1984). On the other hand, the limitation of a vast majority of these observations is the place specific nature inherent in the experimental methodology, i.e., most of these measurements are made at the more accessible basal end of the cochlea and hence conclusions can only be drawn about the developmental status of the high-frequency end of the cochlea.

In the gerbil, the cochlear microphonic at the mid-basal turn of the cochlea is adult-like by 30 days after birth (dab) (Harris and Dallos, 1984) but the summating potential (Yancey and Dallos, 1985) and the endocochlear potential are not (McGuirt *et al.*, 1995). Even more basal in the cochlea (1.2 mm from the base; CF=34−37 kHz), the magnitude of basilar membrane vibration at 20 dab is similar to the postmortem adult (Overstreet *et al.*, 2002). However, the phase response of the basilar membrane response lags in development and is shallower than the adult post-mortem response, showing a total phase accumulation of 0.5 octaves compared to the 1.5 octave total accumulation seen in the post-mortem adult. This limited phase accumulation in the neonate gerbil is consistent with the lack of an adult-like traveling wave and suggests residual immaturities in the physical properties of the basilar membrane.

Cochlear amplifier function, as measured by the DPOAE I/O function, is not mature in gerbils across the entire frequency range until approximately 30 dab (Mills and Rubel, 1996). Mills and Rubel have hypothesized that these changes in cochlear function with age are driven by passive changes in basilar membrane mechanics, rather than maturation of cochlear amplifier function. Consistent with this hypothesis, they describe a shift in the passive cut off frequency, i.e., the

highest frequency at which the basilar membrane can sustain a traveling wave, by approximately two octaves between 14 dab and adulthood. Neural tuning curves recorded from spiral ganglion cells in the midbasal turn of the cochlea appear to be adult like 17 dab (Echteler *et al.*, 1989). The auditory brainstem response (ABR) is first recordable in gerbils at 14 dab but shows continuing changes in latency and amplitude through 40 dab when it becomes fully adult-like (Smith and Kraus, 1987).

In summary, although some aspects of cochlear mechanics and biochemistry appear to be adult-like in the 30-day old gerbil, others, such as the phase response of the basilar membrane and the endocochlear potential are still maturing, albeit at a slower pace compared to the weeks of rapid development preceding this time period. Thus, although the active processes in the cochlea may be mature by 30 dab, immaturity of essential "substrates," such as the correct phase response of the *passive* basilar membrane mechanics, may be preventing the entire system from attaining adult-like functionality (see Overstreet *et al.*, 2002).

The gerbil cochlea around 30 dab may be a valid analog to the human infant around the time of birth. The gerbil has a 25 day gestational period and an onset of hearing at 14 dab (Smith and Kraus, 1987). At 30 dab, (55 days postconceptional age), the gerbil has been hearing for 16 out of 55 days, or 29% of its postconceptional existence. Similarly, a human infant around birth (38–40 weeks), with an onset of hearing at 28 weeks (Krumholz *et al.*, 1985), has been hearing for 10–12 out of 38–40 weeks, or 28% of the postconceptional period. Morey and Carlile (1990) used the period between conception and onset of hearing (the "silent period") in mammals as a foundation to equate the developmental timetables of various species. The gerbil at 30 dab (55 days postconceptional age) and the term-born human [40 weeks postconceptional age (PCA)] are at 141% and 142% of their silent periods, respectively. Thus, both of these approaches suggest that 30 dab in gerbils may be a rough analog to term

(38–40 weeks PCA) in human. Given what is known about gerbil cochlear function at 30 dab, it is possible that outstanding immaturities in physical properties of the basilar membrane could impede adult-like passive vibration along the entire basilar membrane of the human infant around the time of birth.

Similar to the immaturities observed in gerbil, significant OAE immaturities have been consistently observed in human newborns (see Abdala and Keefe, 2006, for a recent summary). The difference in DPOAE level between newborns and young adults (~7 dB) observed in this data set is consistent with reports in the literature (Lafreniere *et al.*, 1991; Bonfils *et al.*, 1992; Lasky *et al.*, 1992; Smurzynski *et al.*, 1993; Abdala, 1996). A significant portion of this difference has recently been attributed to immaturities in the outer and middle ears in the newborn human (Abdala and Keefe, 2006; Keefe and Abdala, 2007). However, discrepancies remain between DPOAE STCs of adults and 6-month old infants, when the middle ear transmission characteristics are predicted to approximate adult-like values (Abdala and Keefe, 2006). Thus, the possibility of residual immaturities in the human infant cochlea cannot be ruled out.

B. DPOAE components

The DPOAE level (L_{dp}) averaged across frequency and several subjects approximates the level of the component from the overlap region (L_{ovlp}) (Brown *et al.*, 1996; Shaffer and Dhar, 2006). Thus the larger L_{dp} in infants is directly related to a larger L_{ovlp}. Fine structure depth, on the other hand, is related directly to the level of the component from the DP CF region (L_{cf}), as evidenced by the disappearance of fine structure upon placing the DP CF region in an audiometric notch (Mauermann *et al.*, 1999b). Therefore the observation in the present study, of greater prevalence of fine structure and greater fine structure depth is linked to a larger L_{cf} in the newborns, as compared to the young adults. Further, assuming the relationship in magnitude between the two components is the same between newborns and adults, fine structure depth will be twice the magnitude of the CF component. This is explained by the fact that to the first order of approximation, fine structure depth is essentially the difference between the in-phase and out-of-phase combinations of the two DPOAE components. In our data, fine structure depth in newborns is greater than adult fine structure by approximately 3 dB, which would translate to an L_{cf} component that is larger by approximately 1.5 dB in this age group.

The observation of a larger L_{ovlp} in the newborns could conceivably be attributed entirely to immaturities in the outer/middle ear system rather than the cochlea. Specifically, a combination of stimulus levels that have been attenuated by an immature middle ear system and a boost to the returning DPOAE energy (due to smaller ear canal area) could account for the higher L_{ovlp} observed. Similarly, lower stimulus levels could also be responsible for a higher L_{cf} in the newborns (Konrad-Martin *et al.*, 2001). However, our data indicate L_{ovlp} is greater in infants than adults by 7 dB whereas L_{cf} is greater by a comparatively smaller amount (as calculated earlier). Given that the overlap and the CF com-

ponents are at the same frequency ($2f_1-f_2$), they should be affected similarly during reverse transmission through the outer and middle ear systems. This suggests that the observed difference in the relative magnitudes of each component in infants versus adults is cochlear in origin. Said another way, bereft of any age-related difference in the generation mechanism of the CF component between newborns and adults, we would expect L_{cf} and L_{ovlp} to be larger by the same amount in the newborns or possibly, the age-related difference to be *greater* for L_{cf}, given the lower effective stimulus levels reaching the newborn cochlea.

Females, whether newborn or young adult, showed larger L_{dp} as well as deeper and more prevalent fine structure. In general, these differences between the two sexes were more apparent in the newborns. Differences in the transmission of sound through the outer and middle ears of the two sexes have been demonstrated in infants (Keefe *et al.*, 2000) and can, at least partially, explain these differences. Male infants showed a trend toward larger reflectance and equivalent volumes than female infants at low frequencies. At frequencies >4 kHz, conductance was also higher in males. These male–female differences, although subtle, suggest anatomical differences in the outer/middle ear structures between sexes even in the newborn period. Other work has suggested sex differences in cochlear physiology, as observed through features of the ABR (Don *et al.*, 1993; Sininger *et al.*, 1998) and OAEs (Burns *et al.*, 1992; McFadden, 1993; McFadden and Mishra, 1993; Bowman *et al.*, 2000). Consistent with anatomical data (Sato *et al.*, 1991), Don and Colleagues (1993) estimated a 13% difference in the length of the cochlea between males and females (males> females) using ABR measures. However, more recent data have shown this to be a gross overestimate, with the measured difference from a relatively large pool of temporal bones showing no significant difference in cochlear length between the sexes (Miller, 2007). Whether all or part of the male–female differences observed in our data are due to differences in transmission characteristics of the outer/middle ear can only be determined through careful juxtaposition of OAE and middle ear function measurements made over wide frequency ranges, an important direction for the future.

An immature middle ear in the infant could very well explain one other striking finding in our data set, the overwhelming prevalence of non-log-sine fine structure patterns. At least one way to arrive at these non-log-sine patterns of fine structure is to "mix" multiple iterations of the DPOAE energy, each additional component arriving in the ear canal with an increasing delay. The reader is directed to Dhar *et al.* (2002) for a detailed theoretical treatment of this phenomenon. Briefly, the impedance mismatch at the boundary between the cochlea and the middle ear acts as a barrier to return a portion of the outgoing DPOAE energy back into the cochlea. This returned energy then essentially becomes the stimulus tone in creating a stimulus frequency OAE generated at the CF region of $2f_1-f_2$. The stimulus frequency OAE (or a portion of it) is in turn returned to the ear canal as a part of the measured DPOAE package. Repeated occurrences of this sequence can lead to the observed non-log-sine patterns of fine structure.

These multiple iterations are relatively rare in the human adult, especially using moderate levels and stimulus frequency ratios around 1.2. When seen in adults, they are often observed at wide stimulus frequency ratios ($\sim$1.3) and low stimulus levels ($\sim$45 dB SPL), indicating increased involvement of the CF region. In the newborn however, the boundary condition at the junction of the cochlea and the middle ear is "exaggerated" due to the immaturities in the middle ear system. This would result in an increased proportion of energy being returned to the cochlea leading to multiple iterations and non-log-sine fine structure. The lowered stimulus levels at the infant cochlea (due to immature middle ear attenuation during forward transmission) may also contribute to the generation of these patterns.

Finally, the unexpected finding of broader fine structure in infants is intriguing. To the best of our knowledge three attempts, similar in many fundamentals, have been made to understand the phenomenon of fine structure in different types of OAEs (Kemp, 1979b; Zwicker and Schloth, 1984; Shera, 2003) and the relationship between OAE and hearing-threshold fine structure (Zwicker and Schloth, 1984). A relationship between the fine structures of DPOAEs, spontaneous OAEs, and estimates of loudness has also been demonstrated (Mauermann *et al.*, 2004). Although, an extended discourse about the intricacies of the previous approaches to fine structure spacing is beyond our scope, we apply the common elements in these models to our own observations. At the very basic level all three models are built on the idea of standing waves or interference patterns between the base and certain points (or regions) along the length of the cochlea. The locations of these regions of resonance along the length of the cochlea are determined primarily by the phase properties of the basilar membrane response and in turn are heavily dependent on the input impedance to the cochlea. In a simplistic view of these models then, fine structure would be more widely spaced if the phase gradient were shallower in infants compared to adults, consistent with the observations of Overstreet *et al.* (2002) in infant gerbils. The middle ear plays an important role in all these models and this role is addressed directly by two of the three authors (Kemp, 1979a, b; Shera, 2003). A change in the input impedance to the cochlea induced by a change in the middle ear would induce a shift in fine structure pattern. However, this shift is predicted to alter the entire pattern of fine structure in frequency, without significantly affecting the spacing between adjacent maxima in DPOAE level (Kemp, 1979a, b; Shera, 2003). Thus, the wider fine structure in infants is not easily explained by immaturities in the conductive pathway and may be suggestive of residual cochlear immaturities in the human newborn, possibly related to passive basilar membrane motion.

The peak of the distribution of fine structure spacing ($f/\Delta f$) in adults is around 16, in agreement with reported spacing of spontaneous OAEs (Shera, 2003) as well as spacing of DPOAE fine structure (Long *et al.*, 1999). In contrast, the distribution of spacing is more dispersed and does not show a clear peak in newborns. This scatter may be indicative of a dynamic system going through maturational change at slightly different rates for each individual infant subject.

In closing, we have demonstrated several differences between DPOAE fine structure in human newborns and young adults. Many of these differences can be explained based on differences in outer and middle ear anatomy and physiology between these age groups. However, we also have demonstrated a striking difference in fine structure spacing between these age groups that may be suggestive of residual cochlear immaturities in the term born human. In particular, differences in fine structure spacing between newborns and adults would be consistent with observed immaturities in passive basilar membrane motion in gerbils, which are the last to mature; well after active cochlear processes have attained adult-like function.

Abdala, C. (**1996**). "Distortion product otoacoustic emission (2f1-f2) amplitude as a function of f2/f1 frequency ratio and primary tone level separation in human adults and neonates," J. Acoust. Soc. Am. **100**, 3726–3740.

Abdala, C. (**1998**). "A developmental study of distortion product otoacoustic emission (2f1-f2) suppression in humans," Hear. Res. **121**, 125–138.

Abdala, C. (**2000**). "Distortion product otoacoustic emission (2f1-f2) amplitude growth in human adults and neonates," J. Acoust. Soc. Am. **107**, 446–456.

Abdala, C. (**2001a**). "DPOAE suppression tuning: Cochlear immaturity in premature neonates or auditory aging in normal-hearing adults?," J. Acoust. Soc. Am. **110**, 3155–3162.

Abdala, C. (**2001b**). "Maturation of the human cochlear amplifier: Distortion product otoacoustic emission suppression tuning curves recorded at low and high primary tone levels," J. Acoust. Soc. Am. **110**, 1465–1476.

Abdala, C. (**2003**). "A longitudinal study of distortion product otoacoustic emission ipsilateral suppression and input/output characteristics in human neonates," J. Acoust. Soc. Am. **114**, 3239–3250.

Abdala, C. (**2004**). "Distortion product otoacoustic emission (2f1-f2) suppression in 3-month-old infants: Evidence for postnatal maturation of human cochlear function?," J. Acoust. Soc. Am. **116**, 3572–3580.

Abdala, C., and Chatterjee, M. (**2003**). "Maturation of cochlear nonlinearity as measured by distortion product otoacoustic emission suppression growth in humans," J. Acoust. Soc. Am. **114**, 932–943.

Abdala, C., and Keefe, D. H. (**2006**). "Effects of middle-ear immaturity on distortion product otoacoustic emission suppression tuning in infant ears," J. Acoust. Soc. Am. **120**, 3832–3842.

Abdala, C., Keefe, D. H., and Oba, S. (**2007**). "Distortion product otoacoustic emission suppression tuning and acoustic admittance in human infants: Birth through 6 months," J. Acoust. Soc. Am. **121**, 3617–3627.

Abdala, C., Sininger, Y. S., Ekelid, M., and Zeng, F. G. (**1996**). "Distortion product otoacoustic emission suppression tuning curves in human adults and neonates," Hear. Res. **98**, 38–53.

ANSI (1996). "Specifications for Audiometers (ANSI S3.6–1996) American National Standards Institute (1996) New York.

Bonfils, P., Avan, P., Francois, M., Trotoux, J., and Narcy, P. (**1992**). "Distortion-product otoacoustic emissions in neonates: Normative data," Acta Oto-Laryngol. **112**, 739–744.

Bowman, D. M., Brown, D. K., and Kimberley, B. P. (**2000**). "An examination of gender differences in DPOAE phase delay measurements in

2200   J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

S. Dhar and C. Abdala: Fine structure comparison

normal-hearing human adults," Hear. Res. **142**, 1–11.

Brown, A. M., Harris, F. P., and Beveridge, H. A. (**1996**). "Two sources of acoustic distortion products from the human cochlea," J. Acoust. Soc. Am. **100**, 3260–3267.

Brown, A. M., Sheppard, S. L., and Russell, P. T. (**1994**). "Acoustic distortion products (ADP) from the ears of term infants and young adults using low stimulus levels," Br. J. Audiol. **28**, 273–280.

Burns, E. M., Arehart, K. H., and Campbell, S. L. (**1992**). "Prevalence of spontaneous otoacoustic emissions in neonates," J. Acoust. Soc. Am. **91**, 1571–1575.

Dhar, S., Long, G. R., Talmadge, C. L., and Tubis, A. (**2005**). "The effect of stimulus-frequency ratio on distortion product otoacoustic emission components," J. Acoust. Soc. Am. **117**, 3766–3776.

Dhar, S., and Shaffer, L. A. (**2004**). "Effects of a suppressor tone on distortion product otoacoustic emissions fine structure: Why a universal suppressor level is not a practical solution to obtaining single-generator DP-grams," Ear Hear. **25**, 573–585.

Dhar, S., Talmadge, C. L., Long, G. R., and Tubis, A. (**2002**). "Multiple internal reflections in the cochlea and their effect on DPOAE fine structure," J. Acoust. Soc. Am. **112**, 2882–2897.

Don, M., Ponton, C. W., Eggermont, J. J., and Masuda, A. (**1993**). "Gender differences in cochlear response time: An explanation for gender amplitude differences in the unmasked auditory brain-stem response," J. Acoust. Soc. Am. **94**, 2135–2148.

Dorn, P. A., Piskorski, P., Gorga, M. P., Neely, S. T., and Keefe, D. H. (**1999**). "Predicting audiometric status from distortion product otoacoustic emissions using multivariate analyses," Ear Hear. **20**, 149–163.

Echteler, S. M., Arjmand, E., and Dallos, P. (**1989**). "Developmental alterations in the frequency map of the mammalian cochlea," Nature (London) **341**, 147–149.

Engdahl, B., and Kemp, D. T. (**1996**). "The effect of noise exposure on the details of distortion product otoacoustic emissions in humans," J. Acoust. Soc. Am. **99**, 1573–1587.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species–29 years later," J. Acoust. Soc. Am.

Harris, D. M., and Dallos, P. (**1984**). "Ontogenetic changes in frequency mapping of a mammalian ear," Science **225**, 741–743.

Heitmann, J., Waldmann, B., Schnitzler, H. U., Plinkert, P. K., and Zenner, H. P. (**1998**). "Suppression of distortion product otoacoustic emissions (DPOAE) near f1-f2 removes {DP}-gram fine structure - Evidence for a secondary generator," J. Acoust. Soc. Am. **103**, 1527–1531.

Keefe, D. H., and Abdala, C. (**2007**). "Theory of forward and reverse middle-ear transmission applied to otoacoustic emissions in infant and adult ears," J. Acoust. Soc. Am. **121**, 978–993.

Keefe, D. H., Folsom, R. C., Gorga, M. P., Vohr, B. R., Bulen, J. C., and Norton, S. J. (**2000**). "Identification of neonatal hearing impairment: Ear-canal measurements of acoustic admittance and reflectance in neonates," Ear Hear. **21**, 443–461.

Kemp, D. T. (**1978**). "Stimulated acoustic emissions from within the human auditory system," J. Acoust. Soc. Am. **64**, 1386–1391.

Kemp, D. T. (**1979a**). "Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea," Arch. Oto-Rhino-Laryngol. **224**, 37–45.

Kemp, D. T. (**1979b**). "The evoked cochlear mechanical response and the auditory microstructure - evidence for a new element in cochlear mechanics," Scand. Audiol. Suppl. 35–47.

Kim, D. O. (**1980**). "Cochlear mechanics: Implications of electrophysiological and acoustical observations," Hear. Res. **2**, 297–317.

Knight, R. D., and Kemp, D. T. (**2000**). "Indications of different distortion product otoacoustic emission mechanisms from a detailed f1,f2 area study," J. Acoust. Soc. Am. **107**, 457–473.

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.

Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (**2001**). "Sources of distortion product otoacoustic emissions revealed by suppression experiments and inverse fast Fourier transforms in normal ears," J. Acoust. Soc. Am. **109**, 2862–2879.

Krumholz, A., Felix, J. K., Goldstein, P. J., and McKenzie, E. (**1985**). "Maturation of the brain-stem auditory evoked potential in premature infants," Electroencephalogr. Clin. Neurophysiol. **62**, 124–134.

Lafreniere, D., Jung, M. D., Smurzynski, J., Leonard, G., Kim, D. O., and Sasek, J. (**1991**). "Distortion-product and click-evoked otoacoustic emissions in healthy newborns," Arch. Otolaryngol. Head Neck Surg. **117**, 1382–1389.

Lasky, R., Perlman, J., and Hecox, K. (**1992**). "Distortion-product otoacoustic emissions in human newborns and adults," Ear Hear. **13**, 430–441.

Lasky, R. E. (**1998**). "Distortion product otoacoustic emissions in human newborns and adults. I. Frequency effects," J. Acoust. Soc. Am. **103**, 981–991.

Long, G. R., Lee, J., and Talmadge, C. (**2004**). "Using sweeping tones to evaluate DPOAE fine structure," In Abstracts of the 27th Midwinter Research Meeting of the ARO, 1076 (Association for Research in Otolaryngology, Des Moines, IA).

Long, G. R., Shaffer, L. A., Dhar, S., and Talmadge, C. L. (**1999**). "Cross species comparison of otoacoustic fine-structure," *Recent Developments in Auditory Mechanics* (World Scientific, Singapore).

Mauermann, M., Long, G. R., and Kollmeier, B. (**2004**). "Fine structure of hearing threshold and loudness perception," J. Acoust. Soc. Am. **116**, 1066–1080.

Mauermann, M., Uppenkamp, S., van Hengel, P. W., and Kollmeier, B. (**1999a**). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. I. Fine structure and higher-order DPOAE as a function of the frequency ratio f2/f1," J. Acoust. Soc. Am. **106**, 3473–3483.

Mauermann, M., Uppenkamp, S., van Hengel, P. W., and Kollmeier, B. (**1999b**). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. II. Fine structure for different shapes of cochlear hearing loss," J. Acoust. Soc. Am. **106**, 3484–3491.

McFadden, D. (**1993**). "A masculinizing effect on the auditory systems of human females having male co-twins," Proc. Natl. Acad. Sci. U.S.A. **90**, 11900–11904.

McFadden, D., and Mishra, R. (**1993**). "On the relation between hearing sensitivity and otoacoustic emissions," Hear. Res. **71**, 208–213.

McGuirt, J. P., Schmiedt, R. A., and Schulte, B. A. (**1995**). "Development of cochlear potentials in the neonatal gerbil," Hear. Res. **84**, 52–60.

Miller, J. D. (**2007**). "Sex differences in the length of the organ of Corti in humans," J. Acoust. Soc. Am. **121**, EL151–EL155.

Mills, D. M. (**2004**). "Relationship of neural and otoacoustic emission thresholds during endocochlear potential development in the gerbil," J. Acoust. Soc. Am. **116**, 1035–1043.

Mills, D. M., and Rubel, E. W. (**1994**). "Variation of distortion product otoacoustic emissions with furosemide injection," Hear. Res. **77**, 183–199.

Mills, D. M., and Rubel, E. W. (**1996**). "Development of the cochlear amplifier," J. Acoust. Soc. Am. **100**, 428–441.

Morey, A. L., and Carlile, S. (**1990**). "Auditory brainstem of the ferret: Maturation of the brainstem auditory evoked response," Brain Res. **52**, 279–288.

Overstreet, E. H., III, and Ruggero, M. A. (**2002**). "Development of wide-band middle ear transmission in the Mongolian gerbil," J. Acoust. Soc. Am. **111**, 261–270.

Overstreet, E. H., III, Temchin, A. N., and Ruggero, M. A. (**2002**). "Passive basilar membrane vibrations in gerbil neonates: Mechanical bases of cochlear maturation," J. Physiol. (London) **545**, 279–288.

R-Development-Core-Team. (**2006**). *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).

Sato, H., Sando, I., and Takahashi, H. (**1991**). "Sexual dimorphism and development of the human cochlea. Computer 3-D measurement," Acta Oto-Laryngol. **111**, 1037–1040.

Shaffer, L. A., and Dhar, S. (**2006**). "DPOAE component estimates and their relationship to hearing thresholds," J. Am. Acad. Audiol **17**, 279–292.

Shera, C. A. (**2003**). "Mammalian spontaneous otoacoustic emissions are amplitude-stabilized cochlear standing waves," J. Acoust. Soc. Am. **114**, 244–262.

Shera, C. A., and Guinan. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs", J. Acoust. Soc. Am. **105**, 782–798.

Shera, C. A., and ZweigG.. (**1991**). "Reflection of retrograde waves within the cochlea and at the stapes", J. Acoust. Soc. Am. **89**, 1290–1305.

Sininger, Y. S., Cone-Wesson, B., and Abdala, C. (**1998**). "Gender distinctions and lateral asymmetry in the low-level auditory brainstem response of the human neonate," Hear. Res. **126**, 58–66.

Smith, D. I., and Kraus, N. (**1987**). "Postnatal development of the auditory brainstem response (ABR) in the unanesthetized gerbil," Hear. Res. **27**, 157–164.

Smurzynski, J., Jung, M. D., Lafreniere, D., Kim, D. O., Kamath, M. V., Rowe, J. C., Holman, M. C., and Leonard, G. (**1993**). "Distortion-product

and click-evoked otoacoustic emissions of preterm and full-term infants, Ear Hear. **14**, 258–274.

Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (**1999**). "Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **105**, 275–292.

Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). "Model-ing otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517–1543.

Yancey, C., and Dallos, P. (**1985**). "Ontogenic changes in cochlear charac-teristic frequency at a basal turn location as reflected in the summating potential," Hear. Res. **18**, 189–195.

Zwicker, E., and Schloth, E. (**1984**). "Interrelation of different oto-acoustic emissions," J. Acoust. Soc. Am. **75**, 1148–1154.

S. Dhar and C. Abdala: Fine structure comparison

Developing standards for distortion product otoacoustic emission measurements

David M. Mills,[a) M. Patrick Feeney, Eli J. Drake, Richard C. Folsom,
Lianne Sheppard, and Noah S. Seixas
Box 357923 University of Washington, Seattle, Washington 98195–7923

Characteristics of distortion product otoacoustic emission (DPOAE) measurements were investigated by comparing responses from two different emission measurement systems in 40 volunteers (78 ears) and making test–retest measurements of each system in 20 ears. For transformation of results between systems, it was shown that the minimum data set consisted of input–output (growth) functions obtained by stepping stimulus levels across a wide range, for each set of stimulus frequencies (1–8 kHz). Linear transformations were considered which involved either recalibration of the emission amplitude (vertical transformation) or of the stimulus levels (horizontal transformation). Horizontal transformations provided better agreement between growth functions from the two systems. For frequencies 4–8 kHz, the means of the horizontal shifts required ranged from 8 to 14 dB, clearly exceeding test–retest variability. The optimal horizontal transformation was derived and applied uniformly to all emission measurements; correlations *r* =0.81–0.89 were found between transformed emission amplitudes. To minimize the necessity for such transformations and to reduce the variability found both within and between systems, development of standardized equipment and methods is suggested for DPOAE measurements, including: (1) an optimized in-ear probe assembly; (2) use of intensity calibration; and (3) a focus on emission "threshold" measurement and analysis. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2770543]

I. INTRODUCTION

It seems evident that the development of standards—equipment and procedures—for otoacoustic emission measurements is a required step for such measurements to become widely accepted as an effective tool for research and clinical use. The *need* for the development of standards for distortion product otoacoustic emission (DPOAE) measurements is illustrated here by the measurement of emissions in human subjects comparing two different systems, plus a number of test–retest measurements of each system. That improvements in emission testing are required is emphasized not only by the significant differences found in emission amplitudes as measured by the two systems, but also by the discovery of considerable variability in responses both within and between the two systems. A careful examination of the results presented here suggests that a significant portion of the measured variability does not arise in the emissions themselves, but in equipment design, calibration, and measurement procedures, which can be improved.

The need for the measurements reported here arose in a slightly different context, as a consequence of a longitudinal study of hearing involving several hundred construction workers (Seixas *et al.*, 2004, 2005). This study included the annual measurement of audiometric thresholds and

DPOAEs. For the first four years of the study, hearing levels and emissions were measured at training and job sites, with emission measurements employing a portable screening system (Scout, Bio-Logic Systems Corp.). For the five year follow-up study, however, emission measurements were made in a double-walled booth using a custom measurement system connected to an ER-10C probe (Etymotic Research Corp.). In spite of this change, it was considered very important to compare the more recent results with the unique and extremely valuable set of emission measurements made earlier.

Unfortunately, emission measurements do not yet have the same kind of standards that make it possible, for example, to measure audiometric responses with different equipment and obtain valid measurements (ANSI, 2004). Methods and procedures for comparing emission measurements using different systems are certainly not well established, and even the optimal stimulus parameters and methods of analysis are still being explored (e.g., Neely and Gorga, 1998; Kummer *et al.*, 2000; Mills, 2002, 2003). The present results are intended to (1) convey the message that there can be large and significant differences between the results of measurements made by different systems in the same individuals, even using identical procedures and stimulus parameters; (2) evaluate specific methods for determining possible transformations between emission input–output functions measured by different systems; (3) determine accuracy (variability) for the derived optimal transformation when applied uniformly to the results of individual measurements; (4) identify probable sources of variability using test–
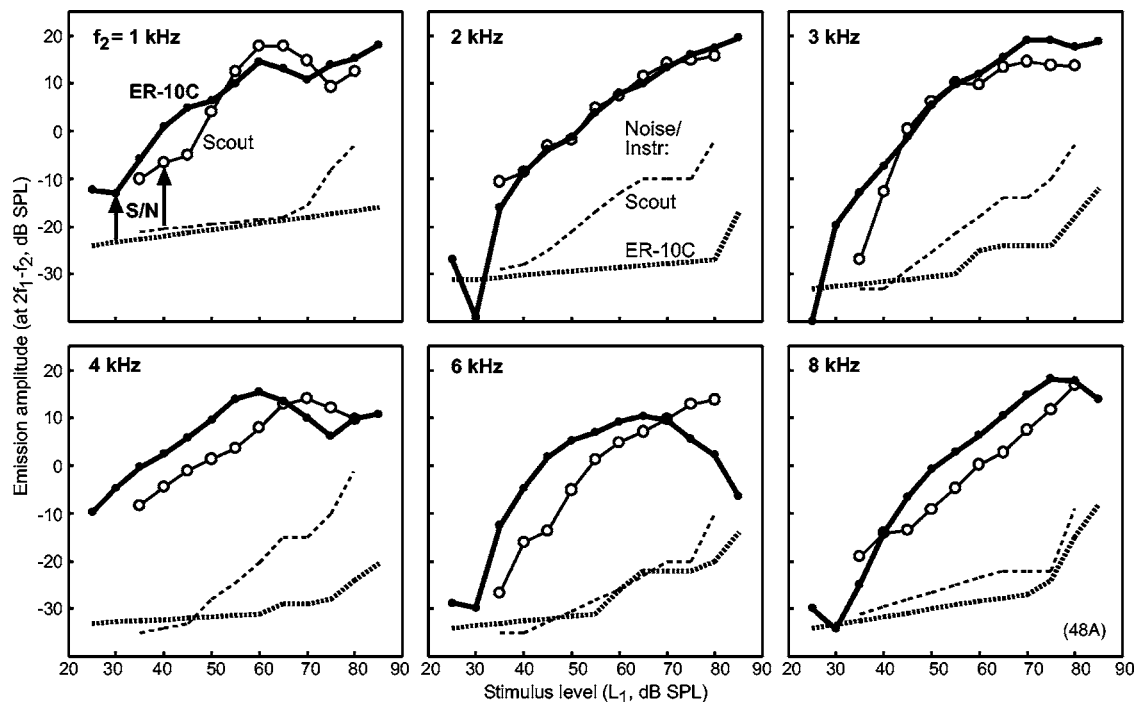
FIG. 1. Typical input-output, or growth functions from one normal-hearing ear. The horizontal axis in each panel is the stimulus level, L_1, in decibel sound pressure level (SPL). The vertical axis represents the amplitude of the bin at the frequency $2f_1$-f_2 in the same units, i.e., the "emission amplitude" reported by each system. The f_2 frequency is noted in the upper left of each panel. Dashed lines are estimates of the intrinsic noise and instrumental distortion levels as a function of stimulus level (see Sec. II A). The solid lines represent measured growth functions for both systems in one ear (see Sec. III C). The amount that the measured amplitude is above the estimated noise/instrumental level is termed the signal-to-noise ratio (S/N) as noted by the arrows. All measured emission amplitudes are shown in these plots irrespective of their S/N.

retest measurements and other results; and (5) discuss possible changes in equipment and procedures intended to eliminate significant measurement differences between different systems and to reduce overall variability in emission measurements.

II. METHODS

A. Estimates of intrinsic noise and instrumental distortion

Prior to measurements in the ear canal, cavity tests were performed by first placing a 1/2 in. reference microphone (Model 2560, Larson-Davis) in one end of a Zwislocki occluded-ear simulator with ear-canal extension (DB-100, Knowles Electronics Corp.). In turn, the foam ear tip/coupler supplied by each manufacturer was placed in the other end, with the distance from the foam tip to the surface of the 1/2 in. microphone chosen to be about the same as the distance from the foam tip to the tympanic membrane (TM) in an ear. For all emission measurements, two stimulus tones (frequencies f_1 and f_2) were applied with the stimulus frequency ratio being $f_2/f_1 = 1.22$. The lower-frequency stimulus level was set to be 10 dB higher than the upper-frequency level ($L_1 = L_2 + 10$ dB). These parameter values were chosen to agree with those in the longitudinal study discussed earlier. At a given set of stimulus frequencies, both stimulus levels were stepped together in 5 dB increments to obtain input–output, or "growth" functions. Comparisons were made of the stimulus levels at f_1 and f_2 as simultaneously measured by the microphone in the probe assembly and by

the 1/2 in. microphone at the other end of the cavity. In addition, the "emission" amplitudes in the cavity, i.e., the contents of the $2f_1 - f_2$ bin as reported by each emission program, were recorded.

B. Measurements in human subjects

Subjects aged 18–45 years were recruited from the university community and paid $25 as an incentive to participate in a two-hour session. Informed consent was obtained, and all materials and methods were approved by the Institutional Review Board at the University of Washington. Subjects were required to pass an otoscopic examination showing no evidence of ear canal or TM abnormality, and to have a normal 226 Hz tympanogram, defined as having a peak tympanometric pressure between −60 and +50 daPa and a peak-compensated static admittance between 0.3 and 1.8 mmho. Two potential subjects were eliminated by this outer/middle ear screening. By design, the study ended when a total of 40 subjects were accepted for testing. Due to subsequent computer malfunction, responses for two ears were lost, leaving 78 ears total.

All emission measurements involved placing into the ear canal the appropriate foam ear tip supplied by the manufacturers (both nominally 1/2 in. diameter uncompressed). Complete growth functions were obtained for each system for six frequencies, $f_2 = 1$, 2, 3, 4, 6, and 8 kHz. The range of levels measured varied by system to match the ranges which had been employed in the longitudinal study, but for each system the same range was employed across frequency and subjects (see Fig. 1).

Mills *et al.*: Developing standards for emission measurements

One of the two systems used for all measurements was the portable system employed for the first four years of longitudinal measurements (Bio-Logic Scout Plus, version 1.75, with AuDX and speaker modules 580-OAE; Bio-Logic Systems Corp.). This system generated tone pips 0.7 s in length and the resulting sound field was sampled, digitized, and stored in a 1 kbyte buffer. A fast Fourier transform (FFT) was performed on the time sample in the buffer and the results displayed. Until stopping conditions were met, a sequence of 0.7 s pips was presented and the FFTs of each were averaged. For the main comparisons (78 ears), stopping conditions were set with user-input parameters identical to those employed in the first part of the longitudinal study (Seixas *et al.*, 2004, 2005). In this mode, averaging times varied between 0.7 and 16.4 s; hence this mode is referred to as the Scout "variable time" measurement. In addition, test–retest measurements were made with the averaging time set to 4 s (Scout "fixed time" mode).

In the same subjects, emission measurements were also made by the system employed in the second half of the longitudinal study. This custom system was headed by a commercially available preamplifier and in-ear probe combination (ER-10C, Etymotic Research). A 24-bit A/D board operating at 96 kHz (Delta 44, M-Audio) was used to obtain analog outputs for the stimulus tones and to sample the microphone signal coming from the preamplifier. The amplitudes of the analog outputs coming from the board were scaled by programmable attenuators (PA-5; Tucker-Davis Technologies) before being sent to the ER-10C preamplifier/ coupler. The attenuators were calibrated *in vivo* so that stimulus levels would be within ±1 dB of their desired values, as measured by the microphone in the in-ear probe assembly.

For a given nominal f_2 frequency, the custom system selected precise frequencies f_1 and f_2 so that the total stimulus repeated exactly every 0.17 s. The microphone response was therefore *time* averaged for the desired duration by having each subsequent 0.17 s sample synchronously averaged in a 16 kbyte buffer. During data acquisition, for artifact rejection purposes, the responses for each 1 s block were first stored in a temporary buffer and an FFT run online. If the noise floor in this 1 s segment was unusually high, indicating an artifactual signal due to subject movement or external sound, this 1 s segment was discarded. After rejecting any artifactual responses, the measurement continued automatically until a total of 6 s of artifact-free signal had been time averaged.

In addition to the comparisons between the two systems measured on all 78 ears, a sequence of test–retest measurements for each system was made in one ear of each subject (20 ears total for each system, randomly selected). These measurements included test–retest measurements with and without probe replacement for both systems, plus comparisons between Scout measurements with fixed and variable times without probe replacement.

C. Methods for data analysis

The main task of the data analysis was to calculate the transformations that would best describe the relationship be-tween the growth functions from each system. For the initial calculation of the transformation, emission amplitudes were ignored if they could reasonably be described as due to or affected by noise and instrumental distortion. A minimum signal-to-noise ratio (S/N) was therefore typically set with respect to the noise/instrumental distortion levels (Fig. 1). For simplicity, the term S/N is understood to include instrumental distortion as well as noise. Following the removal of points falling below the designated minimum S/N, points which were *isolated* were also removed. That is, points were removed which had no nearest neighbors (no neighbors at nominal stimulus levels 5 dB higher or lower). The effect of the assumptions employed for data cleanup on the final outcome was evaluated by carrying the S/N criterion as a parameter throughout the analysis.

Another issue for analysis was to determine the type of transformation needed to reconcile the two emission measurements. After all, at least within limits to be determined, the two measurement systems are measuring the same thing, i.e., the emissions coming from an individual ear. Although the inner ear itself is essentially nonlinear, the emission measurement systems are not. The basic relationship between any two measurement systems should represent a linear transformation, at least for tones below stimulus levels which result in significant saturation of the response. Measurements were always well below this level.

The restriction to a linear transformation means that it is permitted to shift the growth function as seen in a log–log plot in either or both dimensions but not to distort it (see Fig. 1 for typical growth function plots). One example of a linear transformation would be to move all points in a given growth function up, i.e., adding a certain amount in decibels independent of stimulus level. This is equivalent to a microphone recalibration with respect to the measured *emission* amplitude at the relevant frequency. A shift of the growth function in a horizontal direction, on the other hand, is equivalent to recalibrating the microphone with respect to its measurement of *stimulus* levels. The question then arises as to whether one should apply a horizontal or vertical transformation or some combination of the two.

Allowing a linear translation in both directions simultaneously can easily be shown to be an unworkable approach because there are insufficient constraints. That is, for two parallel lines in a log–log plot like Fig. 1, there are an infinite number of combinations of horizontal and vertical displacements that will allow one line to be shifted to overlay the other. However, there *is* a unique transformation if it is restricted to either the horizontal or vertical direction.

As noted, a vertical shift in a plot like Fig. 1 is equivalent to a recalibration of the microphone at the emission frequency $(2f_1 - f_2)$. A horizontal shift in the same plot is equivalent to a recalibration of the microphone at the stimulus frequencies (f_1, f_2). Note that these two calibration functions could be different for a given microphone even at identical frequencies, because of differences in the origin of the measured signals. That is, stimulus tones are generally introduced into the ear canal via tubes with ends located close to the probe microphone, whereas the emissions originate by vibrations of the (more distant) TM driven in reverse.

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Mills *et al.*: Developing standards for emission measurements   2205

Whether horizontal or vertical shifts were considered, it was necessary to define a *fitness measure* to compare the two growth functions. That is, a measure was needed that compared two (input–output) *curves*, not just individual points. To do this, a goodness-of-fit measure was defined by the average of the squared vertical differences between the two curves, over the stimulus levels where both curves were well-defined, i.e., where both emission amplitudes exceeded the criterion signal-to-noise ratio (S/N). The optimal fit was determined by the minimum in this measure, and is reported as the root-mean square (RMS) residual between the two curves.

III. RESULTS

A. Cavity measurements

The following results were noted in preliminary cavity tests. (1) For stimulus frequencies $f_2 \leqslant 1$ kHz, the stimulus levels recorded by the ER-10C system were found to be within ± 1 dB of those measured by the 1/2 in. reference microphone, while the Scout system typically reported a stimulus level about 4 dB higher. (2) At higher frequencies, there was an increasing discordance between the values that each in-ear probe assembly reported for the stimulus levels compared to those measured by the 1/2 in. microphone. For $f_2 = 3$ to 10 kHz, the ER-10C probe underestimated the "actual" stimulus level at the 1/2 in. microphone, whereas the Scout system overestimated it. For both systems, the difference reached a maximum at about 6 kHz, and amounted to about 10–12 dB difference for each (in opposite directions). The shape of the responses suggested that the discordance was due to the difference between the amplitude measured by a probe microphone (with the microphone input located close to the speaker output) compared to that measured at the other end of the simulated ear canal *at frequencies where standing wave effects became important* (e.g., Seigel, 1994; Voss and Shera, 2004). These cavity test results correctly predicted the form of the overall relationship later found between the Scout and ER-10C growth functions when used to measure actual emission amplitudes from human subjects, although the magnitude of the difference in horizontal shift was generally smaller for the actual ears. (3) Finally, at the emission frequency $2f_1 - f_2$, the combination of noise and instrumental distortion was generally found to rise with stimulus level. The steepness of the slope depended on the measurement system and the frequency, but the steepness generally increased with increasing stimulus levels (Fig. 1).

The dashed lines in Fig. 1 show the estimates of noise and instrumental distortion from cavity measurements of the two systems. The observed frequency spectra confirm that values at low stimulus levels represent contributions from noise, whereas the steep increases seen at high stimulus levels represent the effects of instrumental distortion. At the lowest stimulus levels and for frequencies $f_2 \geqslant 2$ kHz, the estimated noise level was from −33 to −35 dB sound pressure level (SPL) for both systems.

B. Test–retest measurements in real ears

Before considering comparisons between systems, one must first determine to what extent the differences found might be due to variations that would occur in one system simply taking successive measurements. That is, one needs to determine the variation in emission amplitude seen when either measurement system repeats its measurement (1) without probe replacement or any other obvious change (except for the passage of several minutes) and (2) with the probe being *replaced* between measurements, but, again, with only a short intervening time and no other changes. The first retest measurement gives an upper bound for short-term variability in the emissions themselves, the second test is obviously the most appropriate for real-world applications of emission measurement systems.

Each test–retest measurement was made on a subset of 20 ears, and typical results are summarized in Fig. 2. In all cases, the means of the differences between successive tests using a given system were not significantly different from zero. This result confirms that the first measurement in general had no deleterious or facilitative effect on emission amplitudes, etc., at least not as measured several minutes later by a second measurement conducted exactly the same way as the first. Therefore, means are not presented here. Rather, the standard deviation (σ) about the mean difference is the statistic used to obtain an estimate of the average size of the test–retest effect.

A typical example of the variation of standard deviation with stimulus level is shown for the Scout system in Fig. 2(A). The results shown for the cases without probe replacement were typical across frequency. There was always a relatively low variance ($\sigma \sim 1$ dB) at high stimulus levels but an increasing variability found at low stimulus levels. The increase cannot be due to noise, because the data for Figs. 2(A) and 2(B) were restricted to measurements with S/N $\geqslant 12$ dB. Instead, the increase in variance seen here suggests that lower emission amplitudes were associated with a higher variability.

There were no significant differences found between either the means of the differences or the variances about the means for measurements comparing fixed to fixed times and variable to fixed times for the Scout. The conclusion is that the change from variable to fixed averaging time will have no practical effect on the comparison of the results from the earlier Scout measurements (using variable times) to those with the ER-10C (using fixed times).

Test–retest measurements for the ER-10C were made for fixed averaging times (6 s) for cases both with and without probe replacement [Fig. 2(B)]. In all cases, the variances found without probe replacement did not have any obvious dependence on frequency. However, the variances with probe replacement did. At low frequencies (1–3 kHz) variances with and without probe replacement were about the same [as in the upper panel in Fig. 2(B)]. However, at higher frequencies (4–8 kHz), there were obvious increases in variances with probe replacement compared to those without [as in lower panel of Fig. 2(B)].

Mills *et al.*: Developing standards for emission measurements
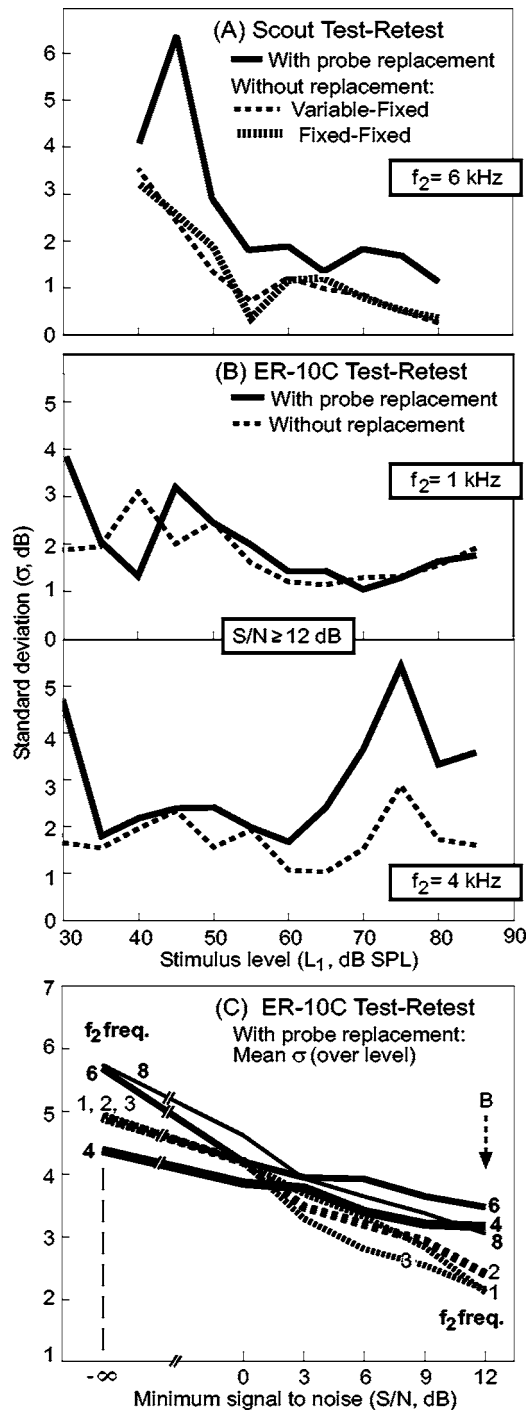
FIG. 2. Summary of test-retest measurements: In (A) and (B), the vertical axis represents the standard deviation (σ) of the difference between test-retest measurements in 20 ears, as a function of the stimulus level, for S/N$\geqslant$12 dB. For (C), these curves have been collapsed by an appropriate average across level, so that the vertical axis is the mean (over stimulus level) of the standard deviation of the test–retest measures with probe replacement. For this calculation, the mean σ was defined as the square root of the mean of the variances (σ^2) summed over stimulus level. The horizontal axis in (C) is the S/N used for the calculations; S/N equal to $-\infty$ simply means that no measured values were excluded. Test conditions for each case are noted on the panels.

Figure 2(C) presents these findings in another way, by summarizing the test–retest results as a function of the S/N (on the horizontal axis). At all frequencies, the typical test–retest variance decreased as the S/N increased. Note that the

continued decrease in variance even for high S/N confirms that low-amplitude emissions typically were associated with higher variance.

The means of the variances summed over both frequency and level support these conclusions. For S/N $\geqslant$9 dB, for example, the mean standard deviations were, in order: Scout test–retest (fixed averaging time) without probe replacement, σ=2.0 dB; ER-10C without replacement, 2.4 dB; ER-10C with replacement, 3.1 dB; and Scout variable with replacement, 3.3 dB. All of these variances were essentially flat across frequency, except for the ER-10C comparison with replacement, which had mean values of σ =2.8 dB over 1–3 kHz and 3.4 dB over 4–8 kHz.

C. Emission growth function comparisons

Examples of emission growth functions measured by both systems are shown by the solid lines in Fig. 1 for one normal-hearing ear. The observed emissions for this ear were generally well above the estimated noise and instrumental distortion levels in both systems. The exceptions for this ear occurred only for the lowest stimulus levels measured, where ER-10C measurements were obviously affected by the noise floor at all frequencies except 4 kHz.

The growth function comparisons in Fig. 1 are illustrative of the overall behavior. The ER-10C and Scout emission amplitudes were generally found to agree fairly well for frequencies f_2=1–3 kHz; this is confirmed below by considering the mean results. However, for frequencies $f_2$$\geqslant$4 kHz the two curves were significantly different, as the examples in Fig. 1 illustrate. Here, the Scout emission amplitudes were typically lower than the ER-10C response at the same stimulus levels. The examples for frequencies of 4 and 6 kHz in Fig. 1 show that it would be incorrect, however, to assume that the observed differences could accurately be reconciled by a simple vertical shift or that they could be properly estimated by a comparison at a single stimulus level.

The main variations seen in observed responses are illustrated by three examples of measured growth functions in Fig. 3. Each panel in the upper row compares original measured growth functions for the two systems, the middle row presents the result of the optimal vertical shift of the Scout response, and the lower row the result of a horizontal shift. The optimal shifts were calculated after data cleaning as described in Sec. II, and the cleaned growth functions are those shown in the lower two rows of panels.

The left column [Fig. 3(A)] illustrates a case with normal emissions in which the horizontal shift clearly gave a much better fit than did the vertical shift. The left middle panel [Fig. 3(A)] illustrates that a vertical shift of +3.8 dB yielded a minimum in the RMS residual, but this residual was 5.3 dB. The lower panel shows that a horizontal shift of −11 dB gave a minimum in the RMS residual which was only 1.2 dB. This numerical result supports the intuitive impression from Fig. 3(A) that the horizontal shift produced a much better agreement between these two growth functions. As will be shown, at each frequency the majority of cases were those in which the horizontal shift gave a better fit. However, this was not true for all ears.
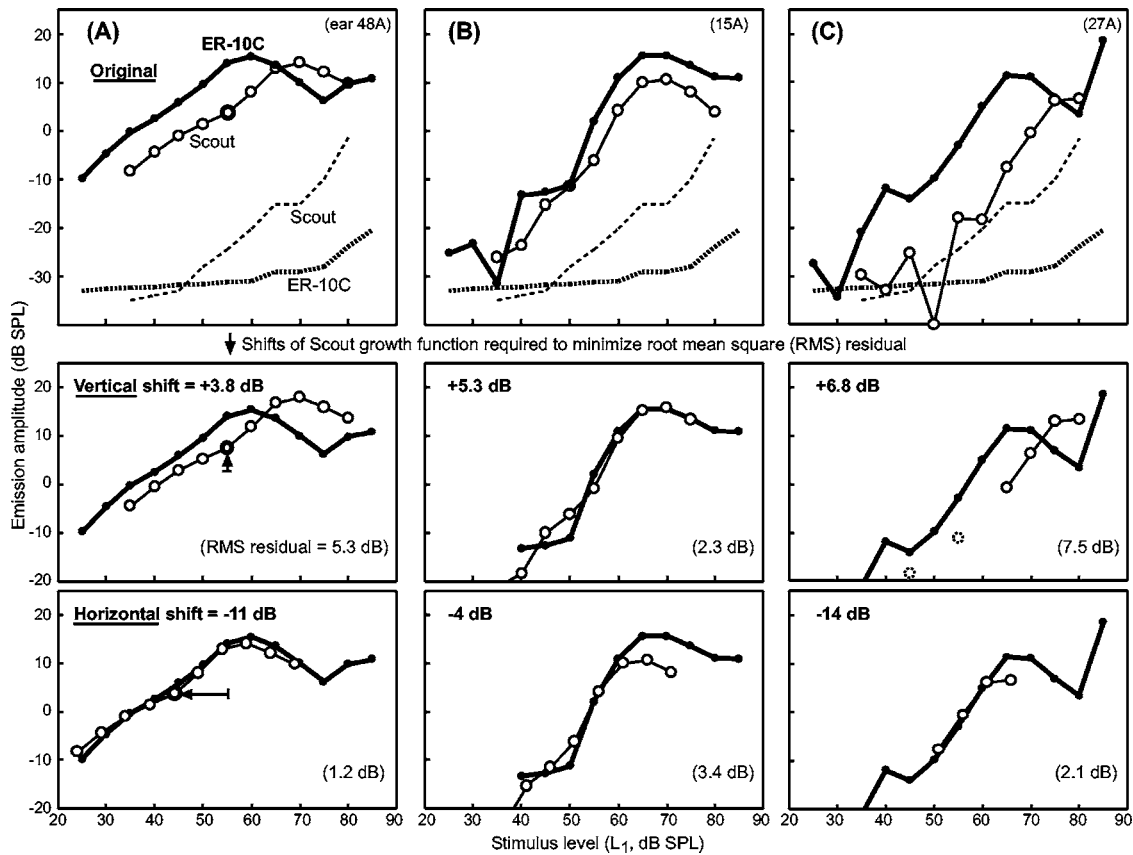
FIG. 3. Examples of shift transformations for three individual ears for $f_2=4$ kHz, one individual in each column. The top panel in each column presents the two measured growth functions. For comparison, noise/instrumental distortion levels (dashed lines) are repeated from Fig. 1. In the top panels, all measured points are shown. Measured points which failed to achieve the minimum distance above the noise/instrumental levels were removed before calculating the shifts in the lower panels. Here, the S/N was set to 6 dB. If, as a result of applying the S/N requirement, there were any "isolated," points, i.e., points without any neighbors with nominal stimulus levels 5 dB away in either direction, these were also removed before the optimal shift was determined. Examples of two such isolated points are shown as dashed circles in the middle panel of column (C). The optimal fit for each shift was then determined by finding the minimum of the root-mean-square (RMS) residual between the two curves as described in Sec. II. The middle panels compare curves after the optimal vertical shift is made, while the bottom panels present results for a horizontal shift. The arrows pointing to the highlighted symbol in the lower left panels indicate the direction and magnitude of the shift for this example. (The two growth functions in the upper left panel (A) are the same as in Fig. 1, repeated for clarity.)

The middle column, Fig. 3(B), provides an example in which the vertical shift gave a (slightly) better fit than the horizontal. For the case shown, the vertical shift of 5.3 dB resulted in a RMS residual of 2.3 dB. This was slightly lower (better) than that for the horizontal shift, which had a RMS residual of 3.4 dB.

Finally, Fig. 3(C) presents typical results for a case in which the emissions were comparatively weak. In such cases, typically many of the Scout emission responses could not be securely differentiated from responses due to noise or instrumental distortion. The lower two panels show only the points remaining after values were removed which were less than 6 dB above the estimated noise/instrumental distortion levels, or which were found to be isolated points after such removal was accomplished. For this case a vertical shift of 6.8 dB did not result in a good fit between the two growth functions, compared to a horizontal shift of −14 dB which did.

A comparison of horizontal and vertical shifts is summarized in Fig. 4. The top panel [Fig. 4(A)] presents the mean shifts as a function of frequency. That is, the vertical axis in Fig. 4(A) represents the mean, taken over the optimal individual shift values, of the population of ears having qualify-ing growth functions, in this case using the minimum S/N of 9 dB. Error bars represent standard deviations. Note that the values for the corresponding standard error of the mean (SEM), because of the large number of ears tested (~78), are typically less than 1 dB. Numerical values for the standard deviation and corresponding SEM are given in Table I for the horizontal shift case in Fig. 4(A).

From Fig. 4(A), for low frequencies ($f_2=1$ and 2 kHz) it is clear that either shift, or no shift, would be adequate to represent the desired transformation. That is, for low frequencies the Scout and ER-10C readings were about the same on average. However, this was certainly not true for frequencies $f_2>3$ kHz. Here, to obtain good agreement between the Scout and ER-10C, either the observed Scout emission amplitudes ($2f_1–f_2$ readings) had to be *increased* by amounts of about 7 dB, or the Scout reported stimulus levels (L_1, L_2) had to be *decreased* by amounts up to 14 dB depending on f_2 frequency. These differences are significant, and well above typical test-retest variances.

The top panel also shows that the variance (scatter) in the numerical values of the shifts derived for individual ears was typically larger for horizontal than for vertical shifts.
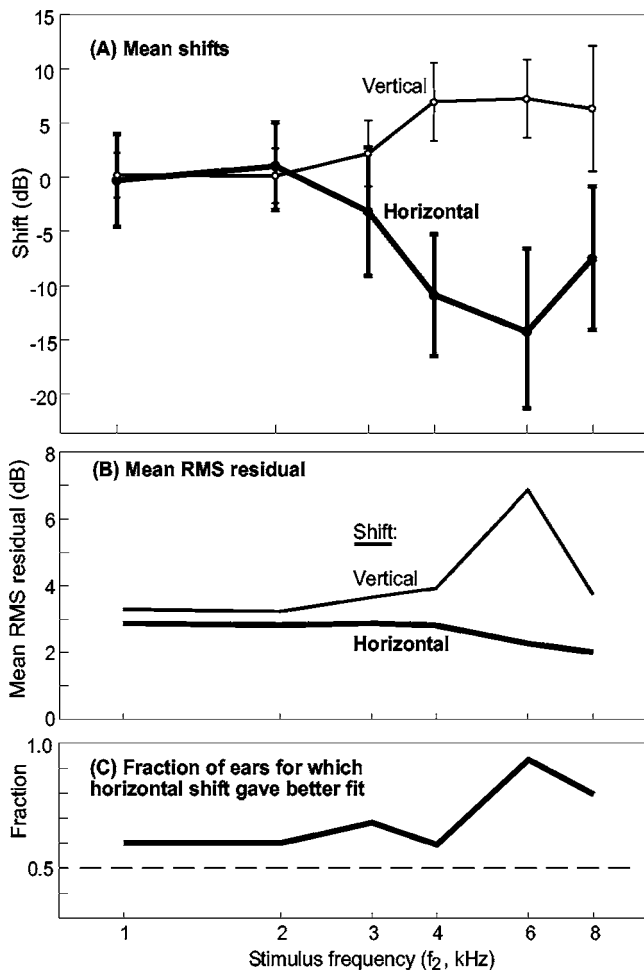
FIG. 4. Comparison of calculated vertical and horizontal shifts for all ears. In all panels, the horizontal axis is the f_2 frequency (kHz). The minimum S/N for this figure was 9 dB. The top panel (A) presents the mean values of the vertical and horizontal shifts, with error bars indicating standard deviations. Panel (B) presents the mean of the RMS residual between the two curves. For this calculation, the mean RMS residual was defined by taking the individual RMS residuals found for each ear, squaring them and taking the mean of the squared values over all ears, then taking the square root. It can be seen that the mean RMS residuals for the horizontal shift were consistently lower than those for the vertical shift, indicating that the horizontal shift typically gave a better fit. This conclusion is reinforced in (C), where the vertical axis represents the *fraction* of individual ears in which the horizontal shift gave a better fit than the vertical shift.

However, the goodness-of-fit between the two growth functions following transformation is correctly estimated using the RMS residual between the two curves (i.e., the RMS residual at the optimal individual shift). The mean values for the calculated RMS residuals are shown in Fig. 4(B). It can be seen, on average, that horizontal shifts gave a better fit than did vertical shifts at all frequencies. To explore this comparison further, Fig. 4(C) presents the *fraction* of cases in which the horizontal shift gave a better fit than the vertical. At all frequencies, over 60% of individual cases had a better fit (lower RMS residual) for the horizontal shift compared to the vertical.

In summation, these results demonstrate that, at all frequencies, a horizontal shift provided a comparatively good fit. In contrast, while the vertical shift gave adequate performance at most frequencies, it failed completely for $f_2=6$ and 8 kHz. In addition, the necessity for at least some horizontal shift at 4–8 kHz was suggested by the cavity measurements.

Because of the results summarized in Fig. 4, the focus henceforth will be on the horizontal shift. As the results in Fig. 4 only considered the case where the criterion S/N =9 dB, it is first of interest to explore how these results varied with S/N. Figures 5 and 6 summarize these findings. In Fig. 5, the horizontal axis is the S/N. The vertical axis is the same quantity as in Fig. 4(B), the mean of the RMS residual over the population of individual ears, with the amount of horizontal shift being optimized for each ear. Across frequency, the goodness-of-fit parameter decreased sharply (the fit improved) between S/N=0 and 3 dB, then improved more slowly with increasing S/N.

In contrast, Fig. 6 shows that there was very little difference in the derived values for the *mean* shift for different S/N ratios. The horizontal axis in Fig. 6 is the stimulus frequency, f_2, and the vertical axis is the mean horizontal shift [as in Fig. 4(A)]. All the curves for the minimum S/N in 3 dB steps from 0 to 12 dB are shown, with the S/N =9 dB case highlighted by use of a dashed line. This case is highlighted because it represents the specific numerical values (in Table I) employed for the "backtransformations" underlying Figs. 7 and 8.

The left-hand columns in Table I list results for the mean horizontal shifts from Fig. 6 for the case with S/N=9 dB. The standard deviations listed are the same as shown for the horizontal transformation case in Fig. 4(A). The number (N) of growth function pairs remaining for comparison after data cleaning are listed in Table I as well as the corresponding values for the SEM. Note that the small differences between the mean shifts in Fig. 6 can be related to the size of the SEM, using the central limit theorem. Finally, definitions for the last two columns of Table I are related to Fig. 7.

Figure 7 presents a check on the accuracy of the derived horizontal transformation in terms of emission amplitudes, obtained by applying a given transformation uniformly to all of the measurements from one system and plotting these values against the measured values for the other system. Figure 7 therefore provides a demonstration of how well the derived transformation, applied uniformly, is able to correct measurements in cases where the emissions were measured with only one system. In the sense that the derivation of the transformation in Figs. 4–6 started from a comparison of measured values from the two systems, Fig. 7 represents a "reverse" transformation.

To derive Fig. 7, the values for the transformation between Scout and ER-10C were taken to be the horizontal shifts of specific values, specifically those listed in left columns of Table I. However, for Fig. 7, *all points that were measured were included* regardless of possible contamination by noise or instrumental distortion. That is, for every measured Scout emission at the stimulus level (L_1, L_2), the corresponding ER-10C stimulus level was found by a horizontal transformation according to Table I. The "measured" responses for the ER-10C at the transformed stimulus levels were then determined by straight-line interpolation between the actual measured values for that individual ear. These are the values plotted on the vertical axes in Fig. 7, against the

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Mills *et al.*: Developing standards for emission measurements   2209

TABLE I. Numerical results for optimal transformation of stimulus levels

	Values for horizontal shift					
	Applicable to Figs. 4 and 6–8				… applicable to Fig. 7.	
Freq (f_2, kHz)	Mean shift (dB)	Std dev (σ, dB)	N	SEM (dB)	Slope (dB/dB)	Correlation, r
1	−0.3	4.3	70	0.5	0.78	0.83
2	+1.0	4.0	70	0.5	0.95	0.89
3	−3.2	5.9	67	0.7	0.90	0.86
4	−10.9	5.6	59	0.7	0.95	0.87
6	−14.3	7.7	62	1.0	0.93	0.87
8	−7.5	6.6	44	1.0	0.97	0.81

observed Scout values on the horizontal axes. In other words, the ER-10C amplitudes at these transformed stimulus levels represent the values, according to the ER-10C system, that the Scout system "should" have given after transformation.

If the transformations in Fig. 7 were perfect, therefore, the best-fit linear regression line would have a slope and *r* value of unity with an intercept through the origin (0,0). In the real situation there is noise and variability, of course, and in some individual cases no horizontal shift of any value could possibly result in a perfect fit [e.g., Fig. 3(B)]. In spite of this, the slopes of the dashed best-fit lines in Fig. 7 were close to unity, intercepts were within 1 dB of the origin (0,0), and *r* values were close to one. Numerical results are summarized in the right two columns of Table I (intercepts not listed). Note again that, while these results use the transformation values originally derived by restriction to relatively high minimum S/N (9 dB), there was *no* restriction on any data in Fig. 7 in terms of the noise floor or instrumental distortion. Therefore, Fig. 7 represents a test of a given transformation when applied to *all* measured data. Obviously, improvement would occur if the Scout measurements were restricted using the noise/instrumental levels shown in Fig. 1. In general, increasing the S/N results in fewer points and less scatter at low emission amplitudes in Fig. 7 (i.e., in the lower left area of each panel), but the slopes and *r* values listed in Table I are changed minimally. Note that the large cluster of

points in the lower left of the plot for 8 kHz is the result of a large number of measurements being at the noise floor in both systems.

It is also important to apply the same transformation as in Fig. 7 to the emission *threshold* (e.g., Mills, 2003). One problem is that the present measurements were not optimized for detection of the emission threshold, e.g., the range for the stimulus levels was limited (for historical reasons). In particular, the Scout measurements went down only to L_1 =35 dB SPL, and this was not always a low enough stimulus level for the emission to fall below the criterion emission threshold amplitude (−10 dB SPL). This was especially a problem for frequencies 1–3 kHz. The existence of these limits might create a bias in the results, because of a tendency to select for cases which happened to have a higher Scout threshold. Nonetheless, Fig. 8 presents a display of all of the emission thresholds that could be estimated from the present data. It can be seen that the measured ER-10C thresholds and the transformed Scout thresholds agreed quite well. The scatter in Fig. 8, which includes observed thresholds at all frequencies, was at least comparable to (if not smaller than) that for the emission amplitudes at individual frequencies (Fig. 7).
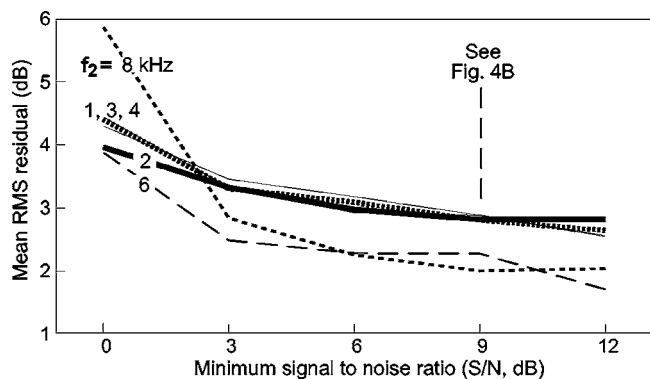
FIG. 5. Summary of the goodness-of-fit determined for *horizontal* shifts as a function of the criterion signal-to-noise ratio (S/N) on the horizontal axis. The vertical axis is the mean (over ears) of the RMS residual calculated as outlined for Fig. 4(B). The stimulus frequency (f_2, kHz) is listed as a parameter. The vertical dashed line indicates that a "cross section" of these responses at the criterion S/N=9 dB can be seen in Fig. 4(B) (horizontal shift case).
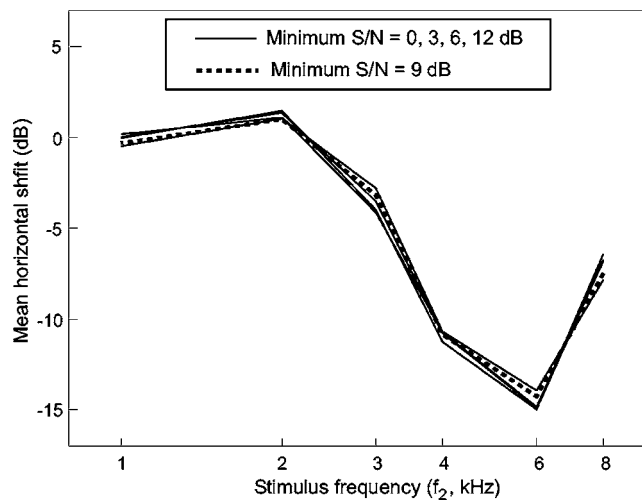
FIG. 6. The mean horizontal shift as a function of stimulus frequency (f_2, kHz) with S/N as a parameter. As there is very little difference between the curves, only the curve for S/N=9 dB is identified separately. The numerical values for this case are reported in Table I, and are also the values used for the calculations for Figs. 7 and 8.
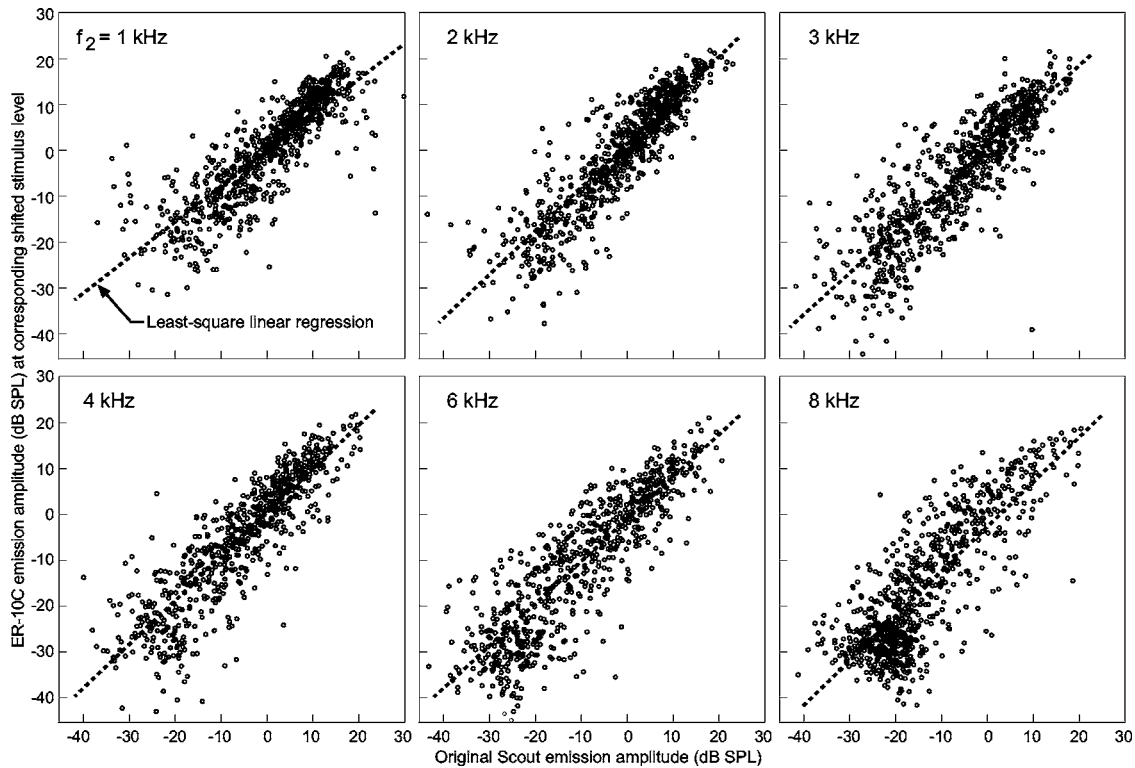
FIG. 7. These plots represent a check on the derivation of the shift transformation by application of one set of shifts to compare the ER-10C emission amplitudes to the Scout measurements. These plots further provide a visual demonstration of the accuracy/repeatability to be expected when mean shifts are applied to individual emission amplitude responses. The shift rules applied were the mean horizontal shifts obtained for the case where the minimum S/N was 9 dB (Table I). The horizontal axis in each panel represents the original Scout emission measurement. This is compared with, on the vertical axis, the ER-10C amplitude for stimulus levels suitably shifted to compare with the Scout (see the text). The dashed lines represent the least-square linear fits to the data points; the relevant numerical results for these regressions are summarized in the right-hand two columns of Table I.

IV. DISCUSSION

A. Relationship of responses measured by different systems

The results of this study demonstrate that, in general, to obtain the transformation between the responses from different measurement systems it is not sufficient to employ the simplest possible transformation, i.e., one cannot use just a recalibration of the microphone response at the emission frequency. In other words, it is not possible in general to correct an individual emission measurement by itself. Rather, at each set of stimulus frequencies, it is necessary to measure growth functions where stimulus levels are stepped across a wide range. This has obvious implications for studies in which emission amplitudes are employed as the only measure for comparison of two systems, e.g., Neely and Gorga (1998). For such studies, measurement of complete growth functions would provide a more accurate estimate of any improvement and would indicate *where* the major differences occurred between the two systems considered, i.e., was it the emission amplitudes or the stimulus levels that caused the main differences in measurements?

To determine the desired transformation between two systems, complete sets of growth functions must be measured comparing the responses of both systems in real ears. Experience from the present study suggests the nature and size of the subject group required. First, one needs only to include subjects likely to have relatively strong emissions;

such a group can easily be obtained by selecting young adult subjects with "excellent" hearing (Mills *et al.*, 2007). The number in the group need not be as large as for the present study, as the main result needed is simply the mean shift as a function of frequency. The typical SEM being 0.5–1 dB with ~70 ears (Table I) suggests that the size of the group could be reduced to as few as ten excellent-hearing subjects (20 ears) *if* having mean shift values accurate to ±1–2 dB were acceptable. As illustrated here, a group of this size would also be generally sufficient for test-retest analysis. These results suggest that useful comparisons could be made between and within different measurement systems with only a modest investment.

B. Sources of variability in emission measurements

There was a relatively large variability found in comparing the amplitude measurements of the two systems in individual subjects (Fig. 7) and in test-retest measurements (Fig. 2). Appreciable variability was present even when emissions were well above the noise floor (e.g., for amplitudes on both axes above −20 dB SPL in Fig. 7). The emission threshold (Fig. 8), which is less sensitive to the noise floor, showed variability comparable to the amplitude measures. Understanding the sources of variability is essential for the development of optimal systems of emission measurement, if the variability is to be reduced.

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Mills *et al.*: Developing standards for emission measurements   2211
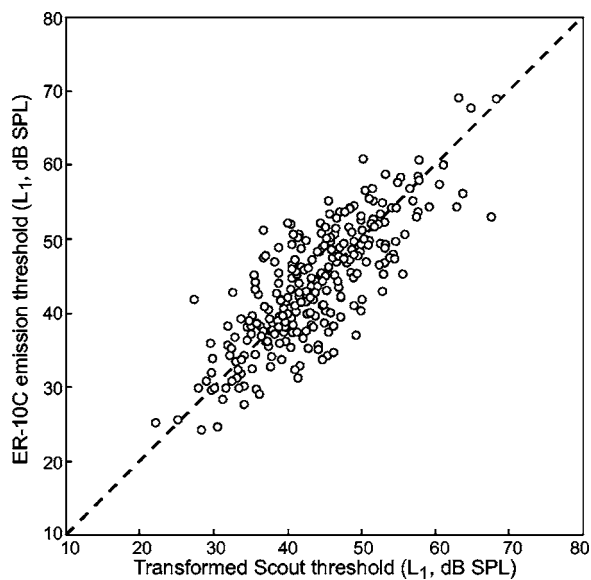
FIG. 8. Similar to Fig. 7, but comparing the Scout emission thresholds (horizontal axis) with the ER-10C emission thresholds (vertical axis). The emission threshold is defined as the stimulus level (L_1) required to achieve a criterion emission amplitude, in this case -10 dB SPL. The observed Scout emission thresholds were transformed by adding the mean shifts in Table I. The dashed line at 45° is for reference. All frequencies were included in this plot, as were all thresholds which could be determined within the stimulus ranges employed for these measurements (see the text).

Of course, there is some variability which can be ascribed to the intrinsic random variation of emission amplitudes. This intrinsic variability is best investigated by test-retest measurements without probe replacement. Logically, the maximum amount of variation which can be ascribed to intrinsic variability is the minimum variability found across systems (at the same parameters). For example, Fig. 2(A) shows that the Scout test–retest variability was given by $\sigma \leq 1$ dB for $L_1 \geq 50$ dB SPL. The ER-10C generally gave slightly higher values in the same range [e.g., dashed lines Fig. 2(B)]. However, the conclusion must be that the intrinsic variation of the emissions themselves can be at most ± 1 dB (for these parameters). Considering all the data, the lowest measured mean variability (over frequency and level) was given by $\sigma = 2.0$ dB. The intrinsic variability of the emissions themselves could be less than this, of course, perhaps ~ 1 dB.

This intrinsic variability can be compared to the total variability in the retest measurements with probe replacement. In the important high frequency region of 4–8 kHz for the ER-10C this was an average $\sigma = 3.4$ dB. This mean value may be slightly misleading in terms of its effect in limiting measurements. The calculated means are dominated by low variability in the least interesting parameter regions, i.e., for high stimulus levels at low frequencies. In contrast, relatively larger variances are found in regions which are of more relevance to researchers and clinicians, i.e., low and moderate stimulus levels at higher frequencies.

Overall, the total variance in emission measurements appears to be 2–3 times the intrinsic variance of the emission amplitudes; there is therefore room for improvement in the variability of emissions as currently measured. A smaller variance would indeed be extremely useful in several situations, e.g., for differential diagnosis of individual cochlear dysfunctions (Mills, 2003) or for DPOAE monitoring of patients. In such situations, changes or differences in individual measurements might have to exceed $\pm 2\sigma$ to provide clinical significance (e.g., Mills, 2006; Mills et al., 2007). Further, the variability in measurement is considerably greater in some parameter regions than in others (e.g., Fig. 2), and there is no way a priori to know where such regions will be when measuring cochleae with different dysfunctions.

Determining the other possible sources of variability in emission measurements may begin with a detailed consideration of the results of test-retest measurements with probe replacement. Such tests were made for the ER-10C system in 20 ears [Figs. 2(B) and 2(C)]. One probable contributor to the observed variability is the inherent variability of the emission amplitude near regions of the growth function which show a "notch," or relative minimum, in the response (e.g., Mills and Rubel, 1994). This is obviously the reason for the strong peak in the mean variability in the 4 kHz response around $L_1 = 75$ dB SPL [Fig. 2(B); compare with the 4 kHz panel in Fig. 1 and with the top panels in Fig. 3]. However, it should be noted that responses at 6 and 8 kHz did not show an obvious increase in variance in regions where notches are often seen. (Reasons for an increased variability near a notch are discussed in the following.)

Overall, there was an observed increase in variability (comparing ER-10C measurements with probe replacement to those without) for frequencies $f_2 = 4$–8 kHz. Considering all of the available information, it seems most likely that this is due to the effect of standing waves in the ear canal (Seigel, 1994; Neely and Gorga, 1998). Evidence here includes the increasing divergence between the measurements made by the two systems as frequency increases from 3 kHz (Fig. 4). A similar divergence was seen in the cavity measurements, with a key observation being that the frequency of divergence decreased with the increase of distance between the probe assembly and simulated TM (data not shown). Note that the larger test–retest variability at high frequencies for measurements in real ears persisted in spite of the fact that the audiologist attempted to insert the probe to the same depth in every measurement, and that the same audiologist was employed for all measurements. It was informally observed that even small changes in the probe position caused by subject movement or by touching the probe could shift the entire high-frequency growth function by several decibels.

Another potential source of variation in emission measurement involves middle ear pressure changes, where relatively small changes in pressure can have large effects on measured emission amplitudes (Zhang and Abbas, 1997). This possibility cannot be completely excluded here because tympanometric responses were measured only once in each ear, at the start of the 2-h testing session. However, in a very similar experiment, the pressure was measured both at the beginning and end of a two-hour session and changes were insignificant for the vast majority of subjects (for 20 of 22 subjects; Mills et al., 2007).

The final factor to consider may be stated as follows. The horizontal conversion was derived assuming that the

stimulus levels, L_1 and L_2, were shifted by the *same* amount, so that the offset between them, 10 dB in this case, would be constant. Such conversion will be inaccurate to the extent that the two stimulus levels (L_1 and L_2) are actually shifted by a *different* amount due to a variation in system response at the frequency f_1 compared to f_2. This difference would change the effective offset between the L_1 and L_2 levels. The extent to which this could potentially be a problem can be estimated from the data in hand, using the rate of change of the horizontal transformation with frequency (Fig. 6). The most rapid change in the derived horizontal transformation was found in the region just below $f_2=4$ kHz. With f_2 equal to 4 kHz and $f_2/f_1=1.22$, the corresponding f_1 is 3.3 kHz. From Fig. 6, the L_2 difference between the two systems at the frequency 4kHz can be estimated to be about 11 dB, compared to the L_1 shift associated with 3.3 kHz which can be estimated to be about 7 dB. This suggests that if one system had an actual L_1, L_2 offset of 10 dB, this could actually have been an offset of either 6 or 14 dB when measured by the other system, depending on which system was taken as baseline.

An examination of Fig. 7 suggests, however, that while differences in the L_1, L_2 offset may be responsible for some of the scatter in the back-transformation, it cannot be the main factor. From Fig. 6, for frequencies of 1 and 2 kHz, the difference in actual offset between the two systems should be near zero, i.e., neither L_1 or L_2 is shifted much. In contrast, the difference in offset should be larger at higher frequencies, with a maximum difference around $f_2=4$ kHz. No such variation is seen in the scatter, which is actually slightly smaller at 4 kHz than at either 1 or 8 kHz (Fig. 7).

C. Standards for equipment and procedures for DPOAE measurements

From the discussion in the previous section, there appears to be several important sources for the test–retest variability found in emission measurements. At the core, there is an apparently irreducible, intrinsic random variability of order ± 1–2 dB in the measurement of emission amplitudes at all frequencies and levels. From the previous discussion, it seems that the variability observed in addition to this core level appears to arise from two main sources: (1) effects due to the presence of notches in the growth function and (2) effects due to standing waves in the ear canal. Potential solutions for each will be examined in turn.

(1) *Notches.* A notch in a growth function is defined as a region with a relative minimum in the emission amplitude. The observed notch can be quite sharp and deep, or more moderate in both qualities. A good example is the relatively consistent notch seen in the 4 kHz growth function at about $L_1=75$–80 dB SPL (Figs. 1 and 3). The consensus is that these notches can be adequately modeled as the addition of two "components" of emission which are of nearly opposite phase (e.g., Mills and Rubel, 1994; Lukashkin and Russell, 1999). The amplitudes of the two components are assumed to vary differently with stimulus level; the notch minimum occurs when the two amplitudes are equal. As the resulting amplitude at and near the notch level is composed of the subtraction of two nearly equal components, the net ampli-

tude is *very* sensitive to small changes in either component. In fact, with stimulus levels held at or near a notch one can observe large changes with repeated measures on a minute-to-minute basis. This appears to be the reason for the strong, broad peak in variance seen in the 4 kHz intrinsic response [lower panel Fig. 2(B)].

One possible solution to this problem might be to simply avoid analyzing emissions measured near notches. One might expect to be able to monitor emissions using a stimulus level well away from a notch region, e.g., using the level $L_1, L_2=60$, 50 dB SPL for the 4 kHz growth function. Unfortunately, this will in general work only while the emission remains normal. When specific lesions were created in animal models, for example, the notch was observed to change its location (Mills, 2003, 2006). In fact, emission *amplitudes* associated with different lesions were found to overlap, the scatter being attributed to the presence of a notch. In contrast, for emission *thresholds*, responses from different lesion types were found to fall in distinct and essentially nonoverlapping regions.

The emission threshold is defined as the stimulus level required to obtain a criterion emission amplitude, typically −10 dB SPL. Appropriate restrictions are required, e.g., to avoid defining a deep notch as a threshold. Determination of an emission threshold typically requires measurement of a growth function. Although this takes more time than a measurement at a single stimulus level, the additional time is not prohibitive. With appropriate software to automate and optimize the procedure for clinical use, a complete set of emission growth functions for six frequencies (1–8 kHz) took about 5 min per ear (Mills *et al.*, 2007). Therefore, one potential solution to the variability due to the presence of notches in emissions is to change the primary measure to be analyzed, from the emission amplitude at a given stimulus level to the emission threshold as defined here (see Mills *et al.*, 2007, re: definitions of emission threshold).

(2) *Standing wave effects.* As frequency increases, standing waves become a problem when the sound pressure in the ear canal becomes significantly different at the TM compared to that at the microphone. This appears to begin to happen at about $f_2=3$ kHz as the responses of the two systems begin to diverge [Fig. 4(A); also see Neely and Gorga, 1998]. Almost certainly, the overall increases in variance for test–retest measures at 4–8 kHz are due to this factor [Fig. 2(C)]. For both of the systems tested here, the stimulus level was calibrated by measuring the sound pressure at the microphone. It has been demonstrated that the stimulus levels at the TM are less sensitive to standing waves if one calibrates instead using the *intensity or power* of the applied stimulus (Keefe *et al.*, 1992; Neely and Gorga, 1998). This requires a separate impedance measurement at the probe position, of course. Pressure calibration was chosen for the present studies so as to match the procedure used in the longitudinal measurements. Further, the results here are applicable to the majority of emission measurements, which continue to use pressure calibration. However, the results summarized here definitely add weight to the proposal that in the long run it may be necessary to change to an intensity calibration protocol.

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Mills *et al.*: Developing standards for emission measurements   2213

The previous sources of variability apply generally to measurements made by any system. The following considerations apply primarily to the comparison of different measurement systems. If two systems use all the same parameters, procedures and methods (as the two tested here did), major differences in their responses *must* have their origin in differences in the acoustic characteristics of their different probe insert assemblies. This includes differences in the acoustic impedance of the probe assembly overall (looking from the ear canal), acoustic/electric characteristics of the probe microphone itself, the location of the speaker tubes relative to other components, and the distance between the probe assembly and the TM. This problem could be minimized if a specific probe configuration could be uniformly adopted for emission measurement systems. Of course, equipment manufacturers may not be enthusiastic about developing a common probe, because this aspect of many systems is proprietary. However, the adoption of a common probe with optimal characteristics and acceptable licensing practices would also provide a level playing field and interchangeability of equipment, in addition to much greater accuracy in comparing the results of different measurement systems.

In this experiment, for example, consider the use of the ER-10C. Although any system built to use the ER-10C probe might well affect the measured noise and instrumental distortion levels, it cannot affect the basic response to emissions above the noise/instrumental distortion floor. Any system using this equipment should then obtain the same basic measurement results as presented here. This has been confirmed by interchanging ER-10C probes among different systems employed at this facility; further, head-to-head cavity testing of different ER-10C probes showed agreement within ±1 dB across frequency (data not shown).

Obviously, before adopting any specific probe configuration, much more work needs to be done in comparing the effects of different probe configurations on the measured emissions, and in comparing the effects of different calibration procedures in combination with different probes. For all potential modifications, attention needs to be paid to improving the repeatability (variability) of the emission responses in test-retest situations. In addition to the adoption of an optimized, standardized in-ear probe assembly, the results obtained here suggest that repeatability may be improved (1) by using intensity calibration, and repeating the calibration frequently, especially during high frequency measurements and (2) choosing an emission threshold measure as the primary approach for measurement and analysis, rather than emission amplitude measures. Emission amplitudes would still be measured, of course, as determination of growth functions over a wide stimulus range is recommended in any case.

It might be argued that it is too early in the development of emission measurements to begin to settle on particular standards. However, at the very least it is not too early to begin to plan experiments and analyze results in such a way as to promote the overall goal of the *eventual* development of such standards. For example, it would be very useful to present adequate test-retest variability measurements to accompany *any* serious proposal regarding equipment or methods. Comparisons of proposed systems with existing technology would also be useful, using an approach like the one illustrated here. Clearly, considerable improvement in emission measurements seems both possible and desirable.

American National Standards Institute (ANSI). (**2004**). "Specification for audiometers," ANSI S3.6-2004, New York.

Keefe, D. H., Ling, R., and Bulen, J. C. (**1992**). "Method to measure acoustic impedance and reflection coefficient," J. Acoust. Soc. Am. **91**, 470–485.

Kummer, P., Janssen, T., Hulin, P., and Arnold, W. (**2000**). "Optimal L_1 - L_2 primary tone level separation remains independent of test frequency in humans," Hear. Res. **146**, 47–56.

Lukashkin, A. N., and Russell, I. J. (**1999**). "Analysis of the f_2-f_1 and $2f_1$-f_2 distortion components generated by the hair cell mechanoelectrical transducer: Dependence on the amplitudes of the primaries and feedback gain," J. Acoust. Soc. Am. **106**, 2661–2668.

Mills, D. M. (**2002**). "Interpretation of standard distortion product otoacoustic emission measurements in light of the complete parametric response," J. Acoust. Soc. Am. **112**, 1545–1560.

Mills, D. M. (**2003**). "Differential responses to acoustic damage and furosemide in auditory brainstem and otoacoustic emission measures," J. Acoust. Soc. Am. **113**, 914–924.

Mills, D. M. (**2006**). "Determining the cause of hearing loss: Differential diagnosis using a comparison of audiometric and otoacoustic emission responses," Ear Hear. **27**, 508–525.

Mills, D. M., Feeney, M. P., and Gates, G. A. (**2007**). "Evaluation of cochlear hearing disorders: Normative distortion product otoacoustic emission measurements," Ear Hear. **28**(6).

Mills, D. M., and Rubel, E. W (**1994**). "Variation of distortion product otoacoustic emissions with furosemide injection," Hear. Res. **77**, 183–199.

Neely, S. T., and Gorga, M. P. (**1998**). "Comparison between intensity and pressure as measures of sound level in the ear canal," J. Acoust. Soc. Am. **104**, 2925–2934.

Seixas, N. S., Goldman, B., Sheppard, L., Neitzel, R., Norton, S., and Kujawa, S. G. (**2005**). "Prospective noise induced changes to hearing among construction industry apprentices," Occup. Environ. Med. **62**, 309–317.

Seixas, N. S., Kujawa, S. G., Norton, S., Sheppard, L., Neitzel, R., and Slee, A. (**2004**). "Predictors of hearing threshold levels and distortion product otoacoustic emissions among noise exposed young adults," Occup. Environ. Med. **61**, 899–907.

Siegel, J. H. (**1994**). "Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes," J. Acoust. Soc. Am. **95**, 2580–2597.

Voss, S. E., and Shera, C. A. (**2004**). "Simultaneous measurement of middle-ear input impedance and forward/reverse transmission in cat," J. Acoust. Soc. Am. **116**, 2187–2198.

Zhang, M., and Abbas, P. J. (**1997**). "Effects of middle ear pressure on otoacoustic emission measures," J. Acoust. Soc. Am. **102**, 1032–1037.

Outer hair cell active force generation in the cochlear environment

Zhijie Liao, Shengran Feng, and Aleksander S. Popel
Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland 21205

William E. Brownell
Bobby A. Alford Department of Otolaryngology-Head and Neck Surgery, Baylor College of Medicine, Houston, Texas 77030

Alexander A. Spector[a]
Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland 21205

Outer hair cells are critical to the amplification and frequency selectivity of the mammalian ear acting via a fine mechanism called the cochlear amplifier, which is especially effective in the high-frequency region of the cochlea. How this mechanism works under physiological conditions and how these cells overcome the viscous (mechanical) and electrical (membrane) filtering has yet to be fully understood. Outer hair cells are electromotile, and they are strategically located in the cochlea to generate an active force amplifying basilar membrane vibration. To investigate the mechanism of this cell's active force production under physiological conditions, a model that takes into account the mechanical, electrical, and mechanoelectrical properties of the cell wall (membrane) and cochlear environment is proposed. It is shown that, despite the mechanical and electrical filtering, the cell is capable of generating a frequency-tuned force with a maximal value of about 40 pN. It is also found that the force per unit basilar membrane displacement stays essentially the same (40 pN/nm) for the entire linear range of the basilar membrane responses, including sound pressure levels close to hearing threshold. Our findings can provide a better understanding of the outer hair cell's role in the cochlear amplifier. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2776154]

I. INTRODUCTION

Outer hair cells are crucial to active hearing providing the amplification and frequency selectivity of the mammalian ear. Such properties of the ear are associated with the active forces produced by outer hair cells. These cells are strategically located in the organ of Corti to pump an additional energy into vibrating cochlear components. The outer hair cell is electromotile (Brownell *et al.*, 1985, Brownell *et al.*, 2001, for review): if the cell is free, it changes the length, and if the cell is constrained, it generates a force. The active properties of the cochlea are particularly pronounced in its high-frequency region. Thus in order to support the mechanism of amplification, the outer hair cell active force has to be significant up to frequencies of tens of kilohertz. In the first proposal for the active cochlea, Gold (1948) foresaw an internal source of additional energy to overcome the high-frequency viscous losses (mechanical filtering) accompanying vibration of the cochlear components. Later experiments (Frank *et al.*, 1999) showed that the outer hair cell does have such capability: Isolated cells driven by ac electric fields in the microchamber generate almost constant forces at frequencies of several tens of kilohertz. The mechanism of outer hair cell motility is associated with the membrane pro-

tein prestin that transfers an electric charge and undergoes conformational changes (Zheng *et al.*, 2000; Liberman *et al.*, 2002; He *et al.*, 2006, for review). While this protein (motor) must have a frequency limit (Gale and Ashmore, 1997), the experiment of Frank *et al.* (1999) shows that such limit is probably large enough to generate forces in the acoustic range.

In contrast to the conditions in the microchamber, the generation of the outer hair cell active force *in vivo* is associated with ac transmembrane potentials that are themselves frequency dependent. Russell *et al.* (1986) presented intracellular receptor potentials in outer hair cells and demonstrated their frequency filtering. Russell and Kössl (1992) presented the outer hair cell receptor potentials measured inside and outside the cell. In particular, they showed that the intracellular potentials are frequency tuned and proportional to the sound pressure level up to 90 dB. The authors also found that these potentials reach 0.2 mV for the sound pressure level equal to 40 dB. The actual potentials in the Russell and Kössl (1992) experiment were probably larger than the reported potentials because of an additional filtering of the recording system. Later, Preyer *et al.* (1996) presented measurements of the receptor potentials in isolated outer hair cells mechanically stimulated via their hair bundles. The authors showed a roll-off of the cell receptor potential beyond a few hundred hertz.

---

[a] Electronic mail: aspector@jhu.edu

In outer hair cells, the transducer current caused by the inclination of the stereocilia bundle and opening of the transduction channels splits into resistive and capacitive components. The latter component increases with the frequency increase resulting in the membrane filtering of the membrane potential. The corresponding frequency roll-off was estimated on the basis of the corresponding RC circuit, including the resistive and capacitive properties of the stereocilia and outer hair cell membrane (Housley and Ashmore, 1992; Santos-Sacchi, 1992). However, the properties of the outer hair cell lateral membrane are also determined by the interaction of ion channels, motor proteins, and the surrounding lipid bilayer, and the subsequent estimates resulted in greater high-frequency receptor potentials (Santos-Sacchi *et al.*, 1998; Spector *et al.*, 2003, 2005; Ospeck *et al.*, 2003)

The fluid environment around the outer hair cells has a finite electrical resistance, and therefore, a portion of the transducer current can flow through the surrounding fluid, changing its potential. For a given outer hair cell, the neighboring outer hair cells contribute to such extracellular potentials, resulting in an additional component of the transmembrane potential important for cell electromotility. Dallos and Evans (1995) noticed that the extracellular potentials do not undergo membrane filtering, and therefore can be effective at high frequencies. Earlier measurements of the potential external to outer hair cells resulted in quite small values (Russell and Kössl, 1992). Recent measurements (Fridberger *et al.*, 2004) showed highly tuned extracellular potentials, although their tuning curve deviates from that of the basilar membrane in the post characteristic frequency (CF) frequency region. The maximal values of the extracellular potential were not insignificant, reaching about 0.17 mV at 18 kHz for 40 dB SPL (Fridberger *et al.*, 2004).

The membrane filtering of the receptor potential has been a critical argument in discussions of the significance of electromotility (Russell and Kössl, 1992). This effect is also important for understanding the contribution of somatic motility versus that of the alternative active mechanism associated with the hair cell's stereocilia bundle (Chan and Hudspeth, 2005; Kennedy *et al.*, 2005; Fettiplace and Hackney, 2006). The high-frequency physiological outer hair cell-related active force was never measured. Earlier estimates (e.g., Russell and Kössl, 1992) were based on a comparison of the cell electromotile length changes corresponding to the measured high-frequency potential with the displacements of the basilar membrane. In the cochlea, however, the cell is not free to change its length, but it is constrained by the relatively stiff basilar membrane/ Deiters' cell and tectorial membrane/reticular lamina complexes.

In this paper, we estimate the outer hair cell active force produced under high-frequency conditions. For that, we construct a model that includes previously developed analyses of high-frequency membrane potentials (Spector *et al.*, 2003, 2005) and the high-frequency cell mechanics (Liao *et al.*, 2005a, b). We also took into account the extracellular potentials provided by the direct experiment (Fridberger *et al.*, 2004). We estimated that, despite the electrical and mechanical filtering in the cochlea, the active force produced by an individual outer hair cell reaches about 40 pN in the high-

frequency region. The active force corresponding to a particular location of outer hair cell inside the cochlea is frequency tuned with the maximal value reached at a point slightly below the local CF. The active force per unit of basilar membrane displacement is sound pressure level invariant, and it is the same for the entire range of the linear basilar membrane mechanics. Thus, the estimated high-frequency outer hair cell's active force can also be significant to low sound pressure levels near the hearing threshold.

II. MODEL

A. Mechanics of the outer hair cell wall

In vivo, the outer hair cell's response to physiological changes in the transmembrane potential is modulated by constraints imposed on the cell by the surrounding solid and fluid components of the cochlea. Our model includes the main cochlear components, such as the basilar membrane, supporting (Deiters') cells, tectorial membrane, reticular lamina, and perilymphatic fluid, that affect the outer hair cell electromotile motion and active force production. Figure 1 sketches the proposed model and explains its correspondence to physiological conditions. In our model, the basilar membrane/Deiters' cell complex and the reticular lamina/tectorial membrane complex are represented by two elastic elements attached to two ends of the cell [Fig. 1(a)]. The external and internal surfaces of the lateral wall of the cell interact, respectively, with the viscous fluids outside and inside the cell. While the outer hair cell wall is a tri-lamina structure, we reduce it to a single layer with effective mechanoelectrical properties. These properties are described by the following constitutive equations:

$$\begin{bmatrix} N_x \\ N_\theta \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{bmatrix} \begin{bmatrix} \epsilon_x \\ \epsilon_\theta \end{bmatrix} + 2\eta \begin{bmatrix} s_x \\ s_\theta \end{bmatrix} + V \begin{bmatrix} e_x \\ e_\theta \end{bmatrix} \quad (1)$$

that relate the internal resultants (forces per unit length) in the cell wall, N_x and N_θ, to the strain, ε_x and ε_θ, and strain rate, s_x and s_θ, components. The last term in the right-hand side of Eq. (1) represents the local active (prestin-related) forces that are generated by changes in the cell transmembrane potential. The subscripts x and θ indicate the axial (along the cell) and circumferential (around the cell) directions, respectively. The mechanical part of Eq. (1) describes a viscoelastic (Voigt-type) model with elastic moduli, C_{ij}, and viscosity η (Evans and Skalak, 1980; Ratnanather *et al.*, 1996; Liao *et al.*, 2005a, b). The elastic part of the model describes an orthotropic two-dimensional (2D) membrane, which is consistent with the contribution of the cell cytoskeleton to the effective properties of the wall (Tolomeo *et al.*, 1996; Tolomeo and Steele, 1995; Spector *et al.*, 2002). We assume here that the C_{ij} moduli are constant, while in more general models (Liao *et al.*, 2005c), they depend on voltage via a voltage-dependent stiffness of the cell (He and Dallos, 1999). The last component (term) in Eq. (1) is of the piezoelectric nature (Mountain and Hubbard, 1994; Tolomeo and Steele, 1995; Spector, 2001; Dong *et al.*, 2002; Deo and Grosh, 2004; Brownell, 2006). Because we model the physiological case of small potentials, this term takes the form of the product of the transmembrane potential, V, and a coeffi-
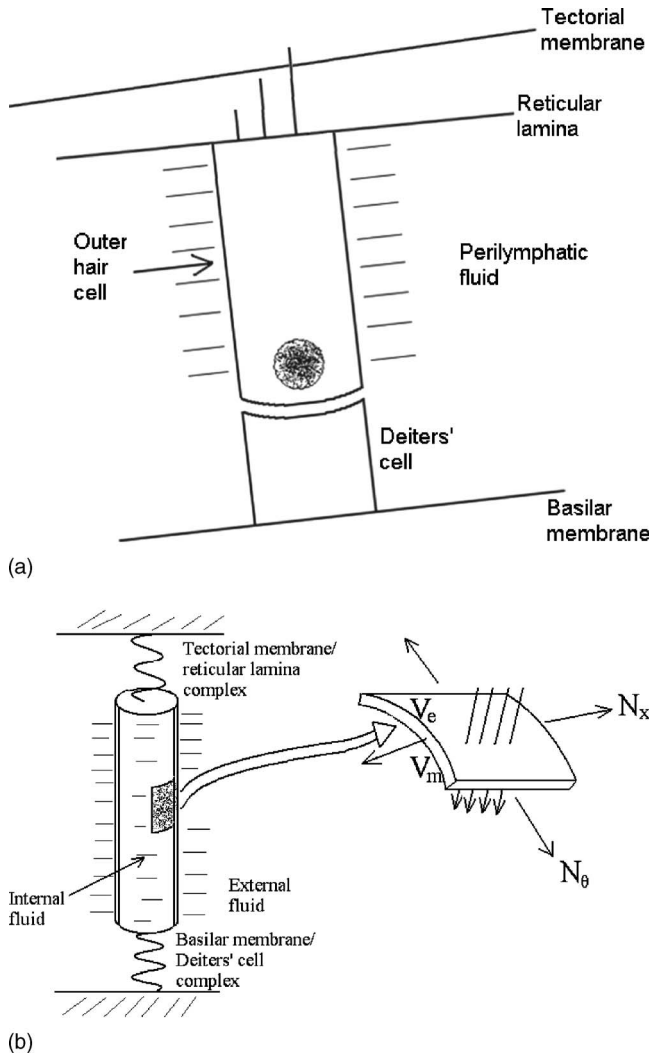
FIG. 1. (a) Sketch of outer hair cell and components of the cochlear environment affecting the cell's active force production, and (b) proposed model of outer hair cell's active force production. Left panel: Cell vibration is constrained by the upper tectorial membrane/reticular lamina and lower basilar membrane/Deiters's cell complexes; the cell wall interacts with the internal (cytosol) and external (perilymph) viscous fluids. Right panel: A representative cut from the cell wall under the action of electric field (arrows) determined by the transmembrane potential V (the difference between the membrane (receptor) potential, V_r, and extracellular potential, V_e). N_x and N_θ are the resultants in the wall.

cient, (e_x, e_θ), associated with the converse piezoelectric effect in the cell membrane. Figure 1(b) shows a representative cut from the cell wall whose properties are described by Eq. (1).

The fluids inside and outside the cell are described by the linearized Navier-Stokes equations (Tolomeo and Steele, 1998):

$$\rho \frac{\partial \mathbf{v}}{\partial t} = -\nabla p + \mu_{i(e)} \nabla^2 \mathbf{v}, \tag{2}$$

where p is the pressure, $\mu_{i(e)}$ is the viscosity of the fluid inside and outside the cell, respectively, $\mathbf{v}$ is the vector of the fluid velocity, and ρ is the fluid density. We assume no-slip boundary conditions where the velocities of fluids and cell wall at interacting surfaces are equal. We neglect the mass of the thin cell wall; thus, in our model, the inertial effects are

associated with the fluid mass vibrating together with the cell.

B. Transmembrane potential

In our model of the outer hair cell active force production, the driving transmembrane potential enters the internal force balance in the cell wall via constitutive Eq. (1). This potential combines two parts, the membrane (receptor) potential, V_r, and extracellular potential, V_e.

The receptor potential is generated by the transducer current in accordance to the properties of the membrane. The outer hair cell membrane comprises lipids, motor proteins, and channels, all of which affect the membrane potential. Such properties are of electrical (resistive and capacitive) and mechanoelectrical nature. The resistive component is associated with voltage gated ionic channels that are mainly located at the basal part of the cell (Housley and Ashmore, 1992; Mammano and Ashmore, 1996; Santos-Sacchi et al., 1997). The capacitive component has two, linear and nonlinear, parts, where the former is associated with the lipid bilayer and the latter results from electric charge transfer by prestin (Housley and Ashmore, 1992; Huang and Santos-Sacchi, 1994; Mammano and Ashmore, 1996). The mechanoelectrical component of the membrane properties is associated with the membrane protein prestin and mechanosensitive channels both distributed along the lateral part of the membrane. Prestin exhibits a converse piezoelectric-like effect by transferring an electric charge across the membrane in response to physiological strain, resulting in a displacement current in the membrane (Gale and Ashmore, 1997; Kakehata and Santos-Sacchi, 1995). The mechanosensitive channels in the outer hair cell membrane were found and described by Ding et al. (1991) and Iwasa et al. (1991) and, more recently, by Rybalchenko and Santos-Sacchi (2003).

Similar to our previous analyses of high-frequency receptor potentials in outer hair cells (Spector et al., 2003, 2005), here we take into account the fine properties of the cell membrane and incorporate them into a model of the receptor potential. This model can be described by

$$V_r = v_r e^{i\omega t} \tag{3}$$

where

$$v_r$$

$$= \frac{(V_{\text{end}} - V_0)G_s^1 d_{\text{BM}} - i\omega \alpha \varepsilon_0 - p^* \varepsilon_0 g_{\text{ch}}(V_0 - V_{\text{ch}})/(1 + i\omega \tau_0)}{G_s^0 + G_m + i\omega C_m + P^*(0,0)G_{\text{ch}}} \tag{4}$$

Here, V_{end} is the endocochlear potential; V_0 is the cell resting potential; G_s^1 and G_s^0 are the total conductance per unit basilar membrane displacement and the "silent" conductance of the transducer channels, respectively; α is the piezoelectric parameter relating the piezoelectric-like displacement current and the strain rate in the membrane; and G_m and C_m are, respectively, the conductance and capacitance of the lateral membrane. The other parameters in Eq. (4) are related to the mechanosensitive channels in the lateral mem-

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Liao et al.: Outer hair cell force generation   2217
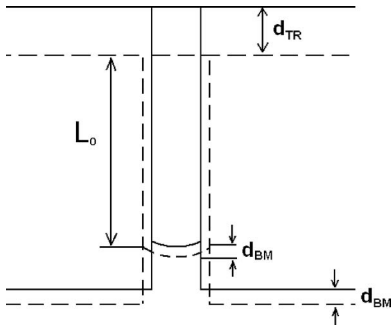
FIG. 2. Sketch of the cell strain due to the relative displacements, d_{TR}, and d_{BM}, of the tectorial membrane/reticular lamina and basilar membrane/Deiters' cell complexes (L_0 is the original length of the cell).

TABLE I. Model parameters of mechanical nature.

Mechanical parameter	Value
Stiffness of the tectorial membrane/reticular lamina complex, S_{TR}	0.2 N/m
Stiffness of the basilar membrane/Deiters's cell complex, S_{BM}	$1-10$ N/m
Cell wall elastic (orthotropic) moduli C_{11}, C_{12}, C_{22}	0.096, 0.16, 0.3 N/m
Perilymth viscosity μ_e	10^{-3} N s/m^2
Cytosol viscosity μ_i	6×10^{-3} N s/m^2
Cell wall (2D) viscosity, η	10^{-7} N s/m
Density of perilymph and cytosol, ρ	10^3 kg/m^3

brane that are assumed to be sensitive to strain and strain rate: p^* is the ratio of the channel's strain and strain rate sensitivity to the strain amplitude ε_0, g_{ch}, and V_{ch} are the total conductance and reverse potential of the channels, respectively. Finally, τ_0 and $P^*(0,0)$ are the parameters of the channel kinetics in a two-state model where the former is the channel typical time and the latter is the steady state value of the probability of the channel being open under the "silent" conditions. The derivation of Eq. (4) is similar to the corresponding analysis in Spector *et al.* (2005). However, here we use the conductance of the transducer channels per unit basilar membrane displacement and the channel's strain/strain rate sensitivity divided by the strain amplitude ε_0. This form of the receptor potential is convenient to further analysis of the outer hair cell active force per unit basilar membrane displacement.

We will now show that the amplitude of the receptor potential is proportional to the amplitude of the basilar membrane displacement, d_{BM}. The first term in the numerator is proportional to d_{BM} [Eq. (4)]. The second and third terms are proportional to the cell strain amplitude, ε_0. The cell ends are assumed to be attached to two vibrating structures, basilar membrane/Deiters' cell and tectorial membrane/reticular lamina complexes. Thus, the cell strain can be defined as the ratio of the difference between the displacements of the two attached structures over the cell length, L_0 (Fig. 2). If we assume that these displacements are inversely proportional to the stiffness of the corresponding complex and also take into account that the basilar membrane/Deiters' cell complex is significantly stiffer (Gummer *et al.*, 1981; Olson and Mountain, 1998; Zwislocki and Ceferatti, 1989; Scherer and Gummer, 2004), the amplitude of the cell strain can be expressed by the following approximate equation:

$$\varepsilon_0 = L_0^{-1} \frac{S_{BM}}{S_{TR}} d_{BM}, \tag{5}$$

where S_{BM} and S_{TR} is the stiffness of the basilar membrane/Deiters' cell and tectorial membrane/reticular lamina complex, respectively. Thus, the whole receptor potential is proportional to the displacement of the basilar membrane, which will be important to our discussion to follow.

We can now consider the second component of the transmembrane potential—the extracellular potential, V_e. Our analysis of this potential will be based on the latest

experimental data of Fridberger *et al.* (2004). According to these data, the extracellular potential can be presented in the following form:

$$V_e = f(\omega, CF) e^{i(\omega t + \varphi)}, \tag{6}$$

where f is the experimental amplitude a "tuning-type" function of frequency, and phase, φ, is about 130° with respect to the basilar membrane displacement. These amplitude and phase were reported by Fridberger *et al.* (2004) in the case of CF=18 kHz. In the following, we present our predictions for a CF different from 18 kHz. To do that, we use the experimental curve for the amplitude f of the extracellular potential from Fridberger *et al.* (2004) and "shift" it along the frequency axis to the corresponding CF. To incorporate the extracellular potentials into our model, we used the points of Fig. 3 from Fridberger *et al.* (2004) with their further analytical approximation. According to its definition, the overall transmembane potential is given by

$$V = V_r - V_e. \tag{7}$$

In this expression of the transmembrane potential, the receptor component, V_r, is given by Eqs. (3) and (4), while its extracellular component, V_e, is described by Eq. (6).

C. Model parameters

The model includes several groups of parameters that characterize the mechanical properties of the cochlear structures, external and internal fluids, and cell wall; electrical properties of the cell stereocilia; and mechanoelectrical properties of the cell membrane. The physiological ranges for these parameters were discussed before (Liao *et al.*, 2005a, b; Spector *et al.*, 2003, 2005). In the following, we present the parameter values chosen for the current modeling: The mechanical characteristics are compiled in Table I and the electrical/mechanoelectrical parameters are presented in Table II.

III. RESULTS

Here, we present the results of our modeling of outer hair cell active force and transmembrane potential under high-frequency conditions. The figures are organized as follows. For both the force and potential, we present their amplitude and phase (with respect to the basilar membrane displacement). The mechanoelectrical properties of the cell

TABLE II. Model parameters of electrical of mechanoelectrical nature.

Electrical (mechanoelectrical) parameter	Value
Endocochlear potential, V_{end}	-80 mV
Cell resting potential, V_0	-70 mV
Steroeocilia silent conductance, G_s^0	0.3 nS
Increment in stereocilia conductance per 1 nm the basilar membrane displacement, G_s^1	1 nS/nm
Piezoelectric coefficient, α	-2.5 nN m/mV
Cell membrane conductance, G_m	40 nS
Cell membrane capacitance C_m	17 pF
Channel reverse potential, V_{ch}	-90 mV
Conductance of mechanosensitive channels, G_{ch}	6 nS
Mechanosensitive channel opening probability under silent conditions, $P(0,0)$	0.5
Complex channel sensitivity, p^*	$-10-0.1\,i\omega$

membrane and the extracellular potentials are critically important in balancing the high-frequency capacitive filtering of the receptor potential (see further discussion in Sec. IV). Because of that, each figure (except the one for force per unit transmembrane potential) includes four graphs where the corresponding characteristic is presented on the basis of electrical (crossed lines), electrical and piezoelectric (dotted lines), electrical, piezoelectric, and channel-related (dashed lines), and all the above-mentioned properties of the membrane in combination with the extracellular potentials (solid lines).

To give a comprehensive analysis of the frequency effect on the active force, we present two sets of figures. First, we show the active force and transmembrane potential versus frequency for constant (frequency-independent) amplitude of the basilar membrane displacement. The basilar membrane is frequency tuned, and its maximal response corresponds to a frequency close to the local CF (e.g., Ruggero, 1992; Robles and Ruggero, 2001). The first set of figures presents the simulation of outer hair cell's responses corresponding to the frequency that causes the maximal (the same throughout the cochlea region under consideration) displacement of the basilar membrane. The results presented in the figures of the second set explicitly take into account the tuning of the basilar membrane: We consider two cells corresponding to two CFs and show their responses to a range of frequencies. For each frequency, the basilar membrane has a different displacement, and the cell has its corresponding response. We assume that the maximal displacements of the basilar membrane for two cells (two CFs) are the same and equal to the displacement considered in the first set of figures. The level of the basilar membrane maximal displacement was chosen at 1 nm, which corresponds to about 30–40 dB SPL (e.g., Nilsen and Russell, 1999; Fridberger et al., 2004). According to Eq. (5) and Table I, this basilar membrane displacement causes a strain of about 0.15% in short outer hair cells located in the high-frequency region of the cochlea.

Figures 3(a) and 3(b) show the amplitude and phase, respectively, of the active force for constant (1 nm) level of the basilar membrane displacement in the frequency range

between 15 and 35 kHz. Figures 3(c) and 3(d) show the amplitude and phase of the transmembrane potential under the same conditions. To illustrate the relative effects of the mechanical and electrical modes of the high-frequency membrane filtering, we present Fig. 4(a) showing the active force under conditions of constant transmembrane potential (force per unit of this potential). To show the effect of the frequency roll-off on the transmembrane potential, Fig. 4(b) presents this potential in the whole frequency range (from low frequencies up to 35 kHz). In Figs. 5(a) and 5(b), we plot the amplitude and phase of the active force for the cell corresponding to CF=18 kHz in the frequency range between 13 and 23 kHz. Figures 5(c) and 5(d) show the amplitude and phase of the transmembrane potential produced by the same cell under the same conditions. Finally, Figs. 6(a) and 6(b) show, respectively, the active force amplitude and its phase, while Figs. 6(c) and 6(d) present; respectively, the transmembrane potential amplitude and its phase for CF=30 kHz and frequency range between 25 and 35 kHz.

IV. DISCUSSION

A. Effect of cell membrane mechanoelectrical properties

Here (Figs. 3, 5, and 6), we concentrate on the effect of the membrane properties on the cell active force. In this regard, Figs. 4(a) and 4(b) provide an important background to look separately at the mechanical and electrical modes of the force modulation. Our data show that the electrical filtering is more significant to the frequency modulation of the cell physiological electromotile response than the mechanical filtering. Indeed, the active force amplitude in Fig. 3(a) monotonically decreases from 52 to 31 pN (about 40%) between frequencies of 15 and 35 kHz. On the other hand, the active force per unit transmembrane potential in Fig. 4(a) is 64 pN/mV at frequency of 15 kHz, and it is 56 pN/mV at frequency of 35 kHz, which corresponds to just about a 12% decrease. Moreover, the latter force has a region of plateau/slight increase from 15 kHz to a frequency of 25 kHz, after which it decreases. Our modeling shows that mechanoelectrical coupling in the outer hair cell membrane (piezoelectricity and mechanosensitive channels) can have a strong effect on the high-frequency active force. All results [e.g., Figs. 3(a), 5(a), and 6(a)] show several-fold greater active force amplitude compared to that predicted by the RC analysis. It is interesting to estimate the effect of the membrane mechanoelectrical properties at different frequencies. The RC analysis predicts a significant decrease of the force with frequency increase. The piezoelectricity contribution stays about the same, and the effect of the mechanosensitive channels decreases with the frequency increase [e.g., Fig. 3(a) and our previous theoretical analysis in Spector et al. (2003, 2005)]. Overall, the relative effect of the mechanoelectrical properties increases with the frequency increase [compare the force amplitudes at 15 kHz with those at 35 kHz in Fig. 3(a)]. This conclusion is true for both the active force and transmembrane potential [Figs. 3(a) and 3(c)].

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Liao et al.: Outer hair cell force generation   2219

FIG. 3. Amplitude and phase of the active force [(a) and (b), respectively] and transmembrane potential [(c) and (d), respectively] in the high-frequency (between 15 and 35 kHz) region of the cochlea. The presented results correspond to a *constant* maximal displacement of the basilar membrane (1 nm). Each characteristic is presented in four versions: crossed lines (RC-approach), dotted lines (electrical+piezoelectric properties of the membrane), dashed lines (electrical+piezoelectric properties of the membrane+mechanosensitive channels), and solid lines (full model, including the extracellular potential).

B. Effect of the extracellular potential

While the extracellular potentials have previously been reported (e.g., Russell and Kössl, 1992), here we use the latest data by Fridberger *et al.* (2004) providing both the potential and basilar membrane displacement under the same conditions. Namely, the data from Figs. 3 and 6 in Fridberger *et al.* (2004) are, respectively, used for the extracellular potential amplitude and phase, and they are incorporated into our model via Eq. (6). In the part of our result presented in Fig. 3, we assumed constant extracellular potential amplitude of about 0.17 mV throughout the frequency range under consideration. The results in Figs. 5 and 6 were obtained by using an approximation of the actual points (Fridberger *et al.*, 2004) of the tuned extracellular potential. From the obtained results, we conclude that the extracellular potential can be a factor in balancing the capacitive filtering of the membrane part of the potential, and it contributes about 30% to the difference between our estimate of the active force and its prediction on the basis of the RC analysis.

C. Basilar membrane tuning

Our model shows that the basilar membrane tuning shapes the frequency dependence of the active force generated by a local outer hair cell. However, the force does not fully follow the frequency response of the basilar membrane displacement. This is because the basilar membrane tuning affects the local active force via the two-component transmembrane potential V [Eq. (7)], where the first of these components, V_r, follows the basilar membrane tuning curve, and the other, V_e, exhibits sharper tuning in the post-CF region. The tuning curves at two different CFs, 18 and 30 kHz, look similar with the maximal values corresponding to frequencies slightly below the corresponding CFs.

However, the maximal value of the active force decreases with increase of the CF: These values are 48 pN [Fig. 5(a)] and 35 pN [Fig. 6(a)] for the CF of 18 and 30 kHz, respectively. This is consistent with the results in Fig. 3(a) that show a frequency decrease in the maximal values of the active forces. The magnitude of the active force drops
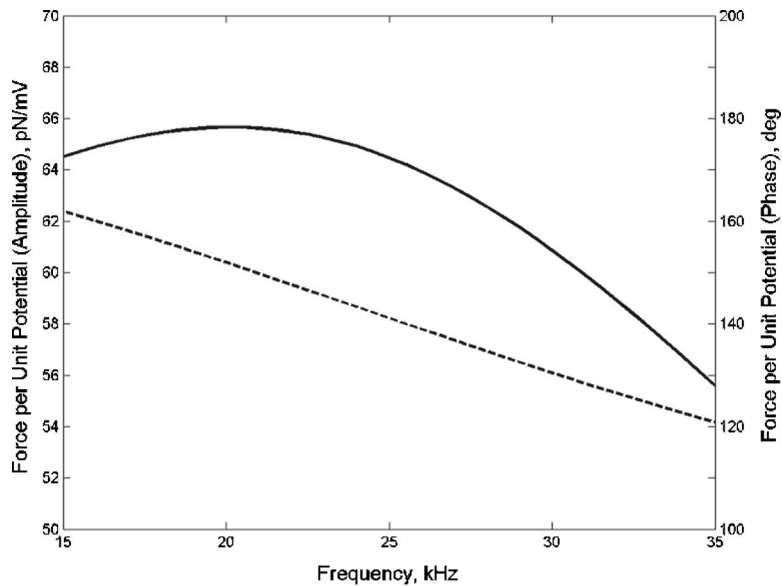
(a)
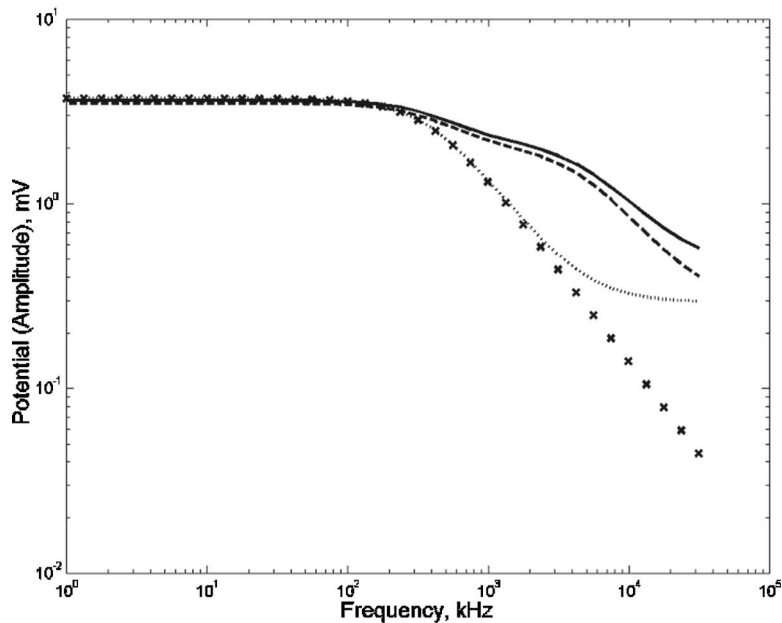
(b)

FIG. 4. (a) Amplitude and phase of the force per unit transmembrane potential in the high-frequency (between 15 and 35 kHz) region of the cochlea, and (b) the transmembrane potential from low to high frequencies. The potential is presented in four versions: crossed lines (RC-approach), dotted lines (electrical+piezoelectric properties of the membrane), dashed lines (electrical +piezoelectric properties of the membrane +mechanosensitive channels), and solid lines (full model, including the extracellular potential).

several-fold outside a band of ±2–3 kHz around the maximal point; and this drop is asymmetric, being steeper in the post-CF region.

The tuning of the active force is significant because it shows that the force production is effective in the narrow frequency range around the local best frequency, and this result is consistent with our understanding of cochlear mechanics.

D. Phase of the active force

Figures 3(b), 4(b), and 5(b) show the frequency dependence of the active force phase relative to the basilar membrane displacement. This phase is 180° at low frequencies, and its frequency modulation has two components. The first is associated with the frequency effect on the force per unit transmembrane potential, and the second is related to the

frequency filtering of the transmembrane potential. The phase of the force per unit transmembrane potential monotonically deviates from 180° with frequency increase [Fig. 4(a)]. The phase of the transmembrane potential slightly changes with frequency increase in all cases under consideration [Figs. 4(d), 5(d), and 6(d)]. As a result of this combination of two frequency effects, the active force phase also deviates from 180° with frequency increase [Figs. 3(b), 4(b), and 5(b)]. Out of four cases of the active force phase considered in Figs. 3(b), 4(b), and 5(b), the largest and smallest deviation from 180° correspond, respectively, to the "RC" and "RC+piezoelectricity" versions of the model. The active force phase corresponding to the model with the mechanosensitive channels and to the full model results in values between the RC and RC+piezoelectricity cases. The effect of the extracellular potential on the active force phase is relatively small, although it produces an interesting local mini-

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Liao *et al.*: Outer hair cell force generation   2221
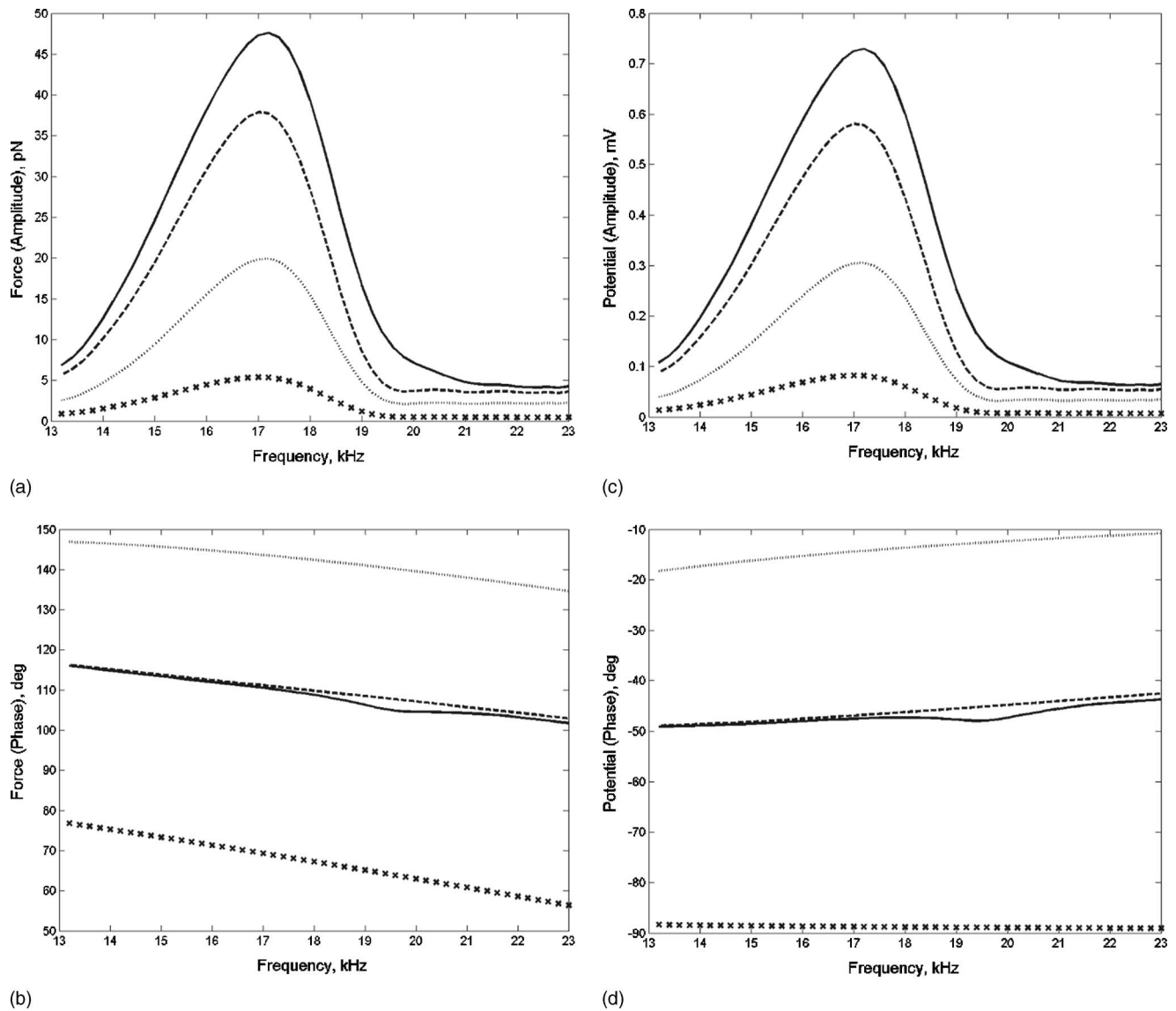
FIG. 5. Effect of the basilar membrane tuning for CF=18 kHz. Amplitude and phase of the active force [(a) and (b), respectively] and transmembrane potential [(c) and (d), respectively] in the high-frequency (between 15 and 35 kHz) region of the cochlea. The presented results correspond to the *local* maximal displacement of the basilar membrane of 1 nm. Each characteristic is presented in four versions: crossed lines (RC-approach), dotted lines (electrical+piezoelectric properties of the membrane), dashed lines (electrical+piezoelectric properties of the membrane+mechanosensitive channels), and solid lines (full model, including the extracellular potential).

mum near CF [solid lines in Figs. 5(b) and 6(b)]. This feature is related to a sharp frequency drop of the extracellular potential in the post-CF region (Fridberger *et al.*, 2004).

E. Outer hair cell transmembrane potential: Model versus experiments

Our predictions of the active force are based on our estimates of the driving transmembrane potential. The obtained results for the transmembrane potential are consistent with a number of qualitative observations in earlier experiments, such as frequency roll-off (Preyer *et al.*, 1996) or frequency tuning of *in vivo* potential (Russell and Kössl, 1992). Due to strong tuning, the magnitude of the receptor potential is highly sensitive to frequencies near the frequency corresponding to the maximal displacement of the basilar membrane (Figs. 5 and 6). This is similar to observations of Russell and Kössl (1992) that showed a significant difference

between the potentials corresponding to frequencies of 19 and 22 kHz recorded at the point of best frequency of 18 kHz. Finally, the maximal values of the intracellular (Russell and Kössl, 1992) and extracellular (Fridberger *et al.*, 2004) potentials corresponding to the basilar membrane displacement of about 1 nm are on the order of magnitude of tenths of a millivolt. This is the same range that is predicted by our model. However, a more comprehensive quantitative comparison between the model transmembrane potential and the experiment is not presently possible due to frequency limitations of the earlier systems [e.g., the time constant of the recording system of Russell and Kössl (1992) was 1–3.5 kHz] and high sensitivity of the results at frequencies close to the local best frequency. Thus, such comparison and, maybe, the corresponding refinement of the model can be made later when newer experimental data become available.
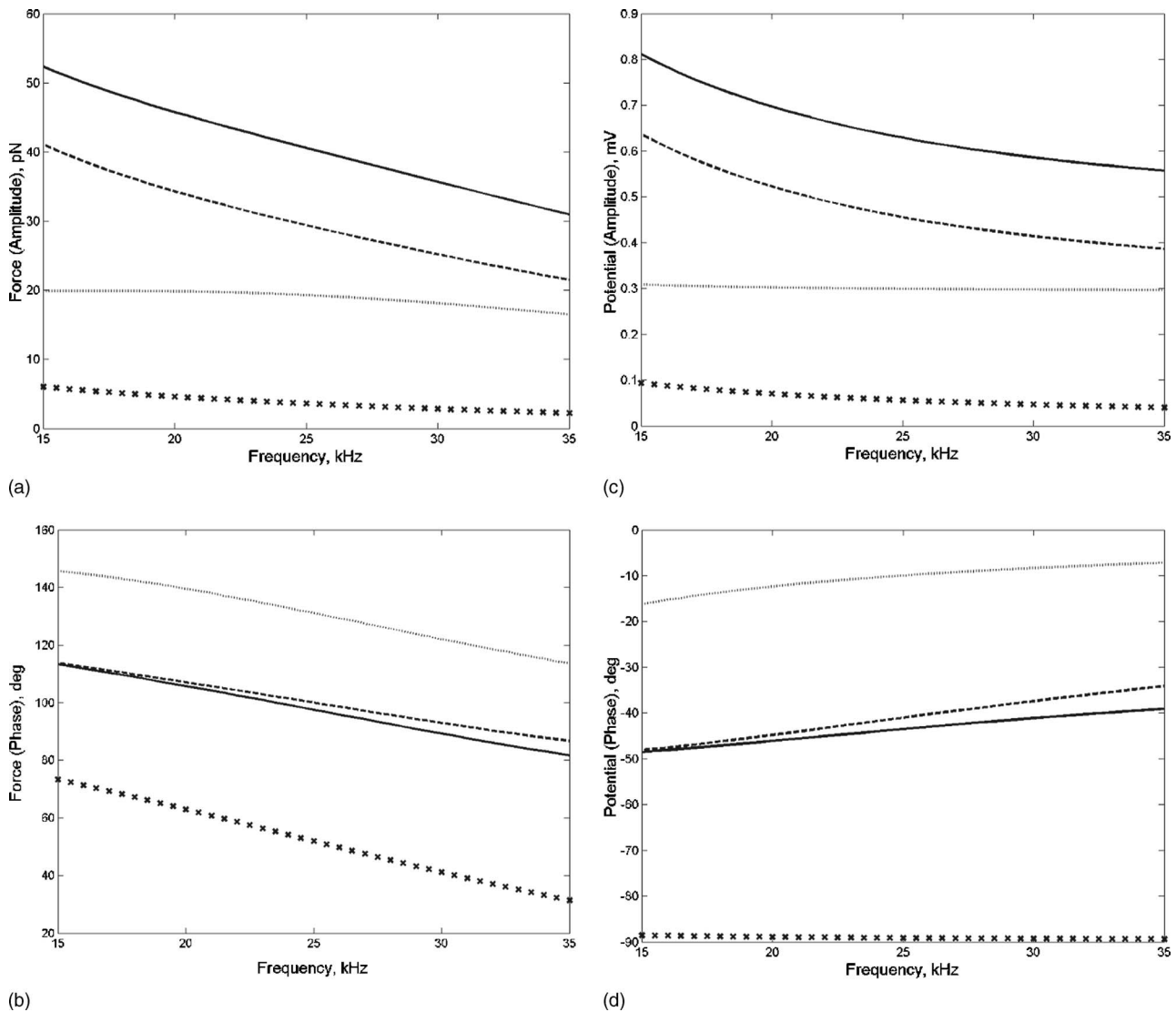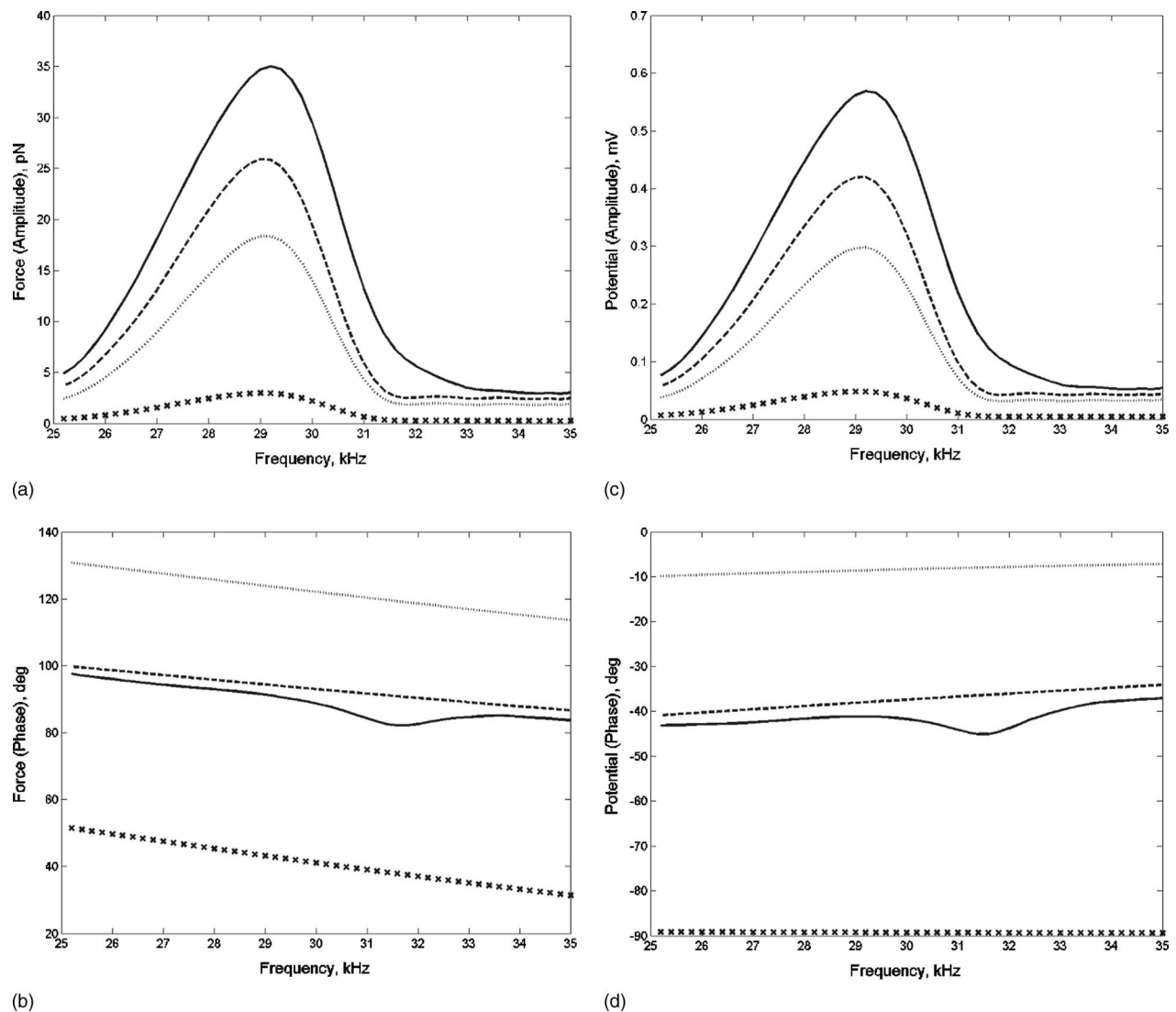
FIG. 6. Effect of the basilar membrane tuning for CF=30 kHz. Amplitude and phase of the active force [(a) and (b), respectively] and transmembrane potential [(c) and (d), respectively] in the high-frequency (between 15 and 35 kHz) region of the cochlea. The presented results correspond to the *local* maximal displacement of the basilar membrane of 1 nm. Each characteristic is presented in four versions: crossed lines (RC-approach), dotted lines (electrical+piezoelectric properties of the membrane), dashed lines (electrical+piezoelectric properties of the membrane+mechanosensitive channels), and solid lines (full model, including the extracellular potential).

F. Significance of the outer hair cell active force to the cochlear amplification

A comprehensive analysis of the significance of the predicted active force will probably require incorporating our individual outer hair cell's model into a dynamic model that includes the elastic, viscous, and inertial properties of the cochlear components. We can, however, make further estimates on the basis of our present analysis.

First, we can estimate the overall active (electromotility related) force involved in the amplification of the basilar membrane. The experiment in guinea pigs shows that the width of the peak of the basilar membrane displacement corresponding to 30–40 dB SPL amplification is about 1 mm (Russell and Nilsen, 1997). This estimate is also consistent with models of the cochlea as a whole (e.g., Geisler, 1998). Assuming that the cell diameter and the distance between neighboring columns of outer hair cells are equal to 10 and 4 μm, respectively, we can estimate that about 70 columns,

or 210 outer hair cells, are involved in the production of this peak. Our estimate of the maximal force is about 35 pN per cell in the region of the CF of 30 kHz. Thus, all these cells can generate a total force of about 7 nN. Note that this may be an overestimate because not all cells produce the force of the same level.

The cochlear amplification is most effective near the hearing threshold, and we can now discuss the significance of the outer hair cell active forces at low sound pressure levels. Our results presented earlier (e.g., Figs. 3, 5, and 6) correspond to the basilar membrane displacements of 1 nm that can be associated with 30–40 dB SPL. For this level of sound intensity as well as below it, the basilar membrane displacement is sound pressure level linear, since the compressive nonlinearity starts at higher sound intensity levels. It can be shown [Eqs. (1) and (5)] that the model outer hair cell's active force is proportional to the basilar membrane displacement. This means that for the linear range of consid-

eration the force is also proportional to the sound pressure level. To use our results in the analysis of lower sound pressure levels, we can introduce the outer hair cell's active force per unit basilar membrane displacement (its maximal amplitude). In the case of 30–40 dB SPL, this characteristic (ratio) will be equal to 40 pN/nm. Thus, in the linear range of the basilar membrane responses, the force per unit basilar displacement will be sound pressure level invariant (stays the same).

In summary, we propose a model of the outer hair cell active force production that takes into account the mechanical, electrical, and mechanoelectrical properties of the cell as well as the effects of the cochlear environment around the cell. The latter includes the viscosity of the perilymph, stiffness of the reticular lamina/tectorial membrane and basilar membrane/Deiters' cell complexes, frequency tuning of the basilar membrane, and acoustic changes in the extracellular electric potential. We investigated both the mechanical and electrical modes of high-frequency filtering of the outer hair cell active force and found that the latter mode is more critical. Our model shows a frequency roll-off and tuning of the active force. We found that despite significant frequency modulation an individual outer hair cell is capable of production of a 40 pN physiological active force under conditions of 30–40 dB SPL. This estimate can be converted to an active force per unit basilar membrane displacement of 40 pN/nm, the same in the whole linear range of the basilar membrane responses. Thus, the estimated active force is also significant at low sound pressure levels close to the hearing threshold.

ACKNOWLEDGMENT

APPENDIX

The main constitutive equations were presented earlier in Sec. II. Here, we describe the major steps of our computational algorithm (see also Liao *et al.*, 2005a, b). First, we specify the basilar membrane displacement, d_{BM}, that determines the transducer current [Eq. (4)] and the cell strain, ε_0. Then, we compute the receptor, extracellular, and transmembrane potentials [Eqs. (4), (6), and (7), respectively]. As a result, we have the electric input via potential V in Eq. (1). After that, the constitutive relation in Eq. (1) can be presented in terms of components, u_x and u_θ, of the elastic displacement of the cell wall

$$
\begin{bmatrix} N_x \\ N_\theta \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{bmatrix} \begin{bmatrix} \dfrac{\partial u_x}{\partial x} \\ \dfrac{u_r}{r_c} \end{bmatrix} + \begin{bmatrix} \eta & -\eta \\ -\eta & \eta \end{bmatrix} \begin{bmatrix} \dfrac{\partial^2 u_x}{\partial x \partial t} \\ \dfrac{\partial u_r}{r_c \partial t} \end{bmatrix}
$$
$$
+ V \begin{bmatrix} e_x \\ e_\theta \end{bmatrix}, \tag{A1}
$$

where r_c is the cell radius. In addition to this relation, we use

the equilibrium equations for the cell wall in the following form:

$$
N_\theta = r_c \sigma_r, \quad N_x = -\int_0^x \sigma_x dx + N_{end} + N_{const}. \tag{A2}
$$

Here, the traction components, σ_r and σ_θ, are determined by the interaction between the cell wall and the fluids via the no-slip conditions. Also, the N_{end} term in Eq. (A2) reflects the end cap interaction with the vibrating fluid and is expressed by

$$
N_{end} = -\frac{k_{end}}{2\pi r_c} u_{end},
$$

where

$$
k_{end} = 16 \left[\frac{i\lambda^2}{6} + (1.32 e^{i0.27\pi} \lambda) \right] \mu_e r_c (i\omega) e^{i\omega t}, \quad \lambda^2 = \frac{\rho \omega r_c^2}{\mu_e}, \tag{A3}
$$

where u_{end} is the axial displacement of the cell end. Finally, N_{constr}—component reflects the cell interaction with the constraints (cochlear structures). Since the basilar membrane/ Deiters' cell complex is much stiffer, we use the fixed-end boundary condition for the lower end of the cell. Thus, N_{const} is defined as

$$
N_{const} = -\frac{S_{TR}}{2\pi r_c} u_{end}. \tag{A4}
$$

The force, $F_a = 2\pi r_c N_{constr}$, is the active force generated by outer hair cell and sensed by the cochlear structures. Thus, Eqs. (A1)–(A4) describe the relationship between the driving transmembrane potential and resulting force produced by the cell.

To obtain a numerical solution, we assume the axisymmetry of the problem with respect to the θ coordinate, and use [following Tolomeo and Steele (1998), Laio *et al.* (2005a, b)] Fourier expansions in terms of the x coordinate. Note that recently Lim and Li (2007) proposed a direct (nonanalytical) finite-difference method to solve outer hair cell-related dynamic problems. Expanding each model function in terms of Fourier series and combining Eqs. (A1)–(A4), the vector of the Fourier coefficients of the elastic displacement, **u**, can be expressed in terms of that of the piezoelectric term, $\boldsymbol{\sigma}_{piez}$, as

$$
\mathbf{u} = [K]^{-1} \boldsymbol{\sigma}_{piez}, \tag{A5}
$$

where the generalized stiffness matrix $[K]$ is determined by the mechanical properties of the cell wall, both fluids, and constraint (reticular lamina/tectorial membrane complex). Computing the displacement vector, we find the end displacement and, finally, the active force produced by the cell.

Brownell, W. E. (**2006**). "The piezoelectric outer hair cell," in *Vertebrate Hair Cells*, edited by R. A. Eatock (Springer, New York), pp. 313–347.

Brownell, W. E., Bader, C. D., Bertrand, D., and de Ribaupierre, Y. (**1985**). "Evoked mechanical responses of isolated cochlear outer hair cells," Science **224**, 194–196.

Brownell, W. E., Spector, A. A., Raphael, R. M., and Popel, A. S. (**2001**). "Micro- and nanomechanics of the cochlear outer hair cell," Annu. Rev. Biomed. Eng. **3**, 169–194.

Chan, C. D., and Hudspeth, A. J. (**2005**). "Mechanical responses of the organ of Corti to acoustic and electrical stimulation in vitro," Biophys. J. **89**, 4382–4395.

Dallos, P., and Evans, B. N. (**1995**). "High frequency motility of outer hair cells and the cochlear amplifier," Science **267**, 2006–2009.

Deo, N., and Grosh, K. (**2004**). "Two-state model for outer hair cell stiffness and motility," Biophys. J. **86**, 3519–3528.

Ding, J. P., Salvi, R. J., and Sachs, F. (**1991**). "Stretch-activated ion channels in guinea pig outer hair cells," Hear. Res. **56**, 19–28.

Dong, X. X., Ospeck, M., and Iwasa, K. H. (**2002**). "Piezoelectric reciprocal relationship of the membrane motor in the cochlear outer hair cell," Biophys. J. **82**, 1254–1259.

Evans, E. A., and Skalak, R. (**1980**). *Mechanics and Thermodynamics of Biomembranes* (CRC Press, Boca Raton, FL).

Fettiplace, R., and Hackney, C. M. (**2006**). "The sensory and motor roles of auditory hair cells," Nat. Rev. Neurosci. **7**, 19–29.

Frank, G., Hemmer, W., and Gummer, A. W. (**1999**). "Limiting dynamics of high-frequency electromechanical transduction of outer hair cells," Proc. Natl. Acad. Sci. U.S.A. **96**, 4420–4425.

Fridberger, A., de Monvel, J. B., Zheng, J., Hu, N., Zou, Y., Ren, T., and Nuttall, F. (**2004**). "Organ of Corti potentials and the motion of the basilar membrane," J. Neurosci. **24**, 10057–10063.

Gale, J. E., and Ashmore, J. F. (**1994**). "Charge displacement induced by rapid stretch in the basolateral membrane of the guinea-pig outer hair cell," Proc. R. Soc. London, Ser. B **255**, 243–249.

Gale, J. E., and Ashmore, J. F. (**1997**). "An intrinsic frequency limit to the cochlear amplifier," Nature (London) **389**, 63–66.

Geisler, C. D. (**1998**). *From Sound to Synapse: Physiology of the Mammalian Ear* (Oxford University Press, Oxford, UK).

Gold, T. (**1948**). "Hearing. II. The physical basis of the action of the cochlea," Proc. R. Soc. London, Ser. B **135**, 492–498.

Gummer, A. W., Johnstone, B. M., and Armstrong, N. J. (**1981**). "Direct measurement of basilar membrane stiffness in the guinea pig," J. Acoust. Soc. Am. **70**, 1298–1309.

He, D. Z. Z., and Dallos, P. (**1999**). "Somatic stiffness of cochlear outer hair cell is voltage-dependent," Proc. Natl. Acad. Sci. U.S.A. **96**, 8223–8228.

He, D. Z. Z., *et al.* (**2006**). "Tuning in to the amazing outer hair cell: Membrane wizardry with a twist and shout," J. Membr. Biol. **209**, 119–134.

Housley, G. D., and Ashmore, J. F. (**1992**). "Ionic currents of outer hair cells isolated from the guinea-pig cochlea," J. Physiol. (London) **448**, 73–98.

Huang, G., and Santos-Sacchi, J. (**1994**). "Motility voltage sensor of the outer hair cell resides within the lateral plasma membrane," Proc. Natl. Acad. Sci. U.S.A. **91**, 12268–12272.

Iwasa, K. H., Li, M. X., Jia, M., and Kachar, B. (**1991**). "Stretch sensitivity of the lateral wall of the auditory outer hair cell from the guinea pig," Neurosci. Lett. **133**, 171–174.

Kakehata, S., and Santos-Sacchi, J. (**1995**). "Membrane tension directly shifts voltage-dependence of outer hair cell motility and associated gating charge," Biophys. J. **68**, 2190–2197.

Kennedy, H. J., Crawford, A. C., and Fettiplace, R. (**2005**). "Force generation by mammalian hair bundles supports a role in cochlea amplification," Nature (London) **433**, 880–883.

Liao, Z., Popel, A. S., Brownell, W. E., and Spector, A. A. (**2005a**). "Modeling high-frequency electromotility of cochlear outer hair cell in microchamber experiment," J. Acoust. Soc. Am. **117**, 2147–2157.

Liao, Z., Popel, A. S., Brownell, W. E., and Spector, A. A. (**2005b**). "High-frequency force generation in the constrained cochlear outer hair cell: A model study," J. Assoc. Res. Otolaryngol. **6**, 378–389.

Liao, Z., Popel, A. S., Brownell, W. E., and Spector, A. A. (**2005c**). "Effect of voltage-dependent membrane properties on active force generation in cochlear outer hair cell," J. Acoust. Soc. Am. **118**, 3737–3746.

Liberman, M. C., Gao, J., He, D. Z., Wu, X., Jia, S., and Zuo, J. (**2002**). "Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier," Nature (London) **419**, 300–304.

Lim, K. M., and Li, H. L. (**2007**). "A two-layer outer hair cell model with orthotropic piezoelectric properties: Correlation of cell resonant frequencies with tuning in the cochlea," J. Biomech. **40**, 1362–1371.

Mammano, F., and Ashmore, J. F. (**1996**). "Differential expression of outer hair cell potassium currents in the isolated cochlea of the guinea-pig," J. Physiol. Paris **496**, 639–646.

Mountain, D. C., and Hubbard, A. E. (**1994**). "A piezoelectric model of

outer hair cell function," J. Acoust. Soc. Am. **95**, 350–354.

Nilsen, K. E., and Russell, I. J. (**1999**). "Timing of cochlear feedback: Spatial and temporal representation of a tone across the basilar membrane," Nat. Neurosci. **2**, 642–648.

Olson, E. S., and Mountain, D. C. (**1991**). "In vivo measurement of basilar membrane stiffness," J. Acoust. Soc. Am. **89**, 1262–1275.

Ospeck, M., Dong, X. X., and Iwasa, K. H. (**2003**). "Limiting frequency of the cochlear amplifier based on electromotility of outer hair cells," Biophys. J. **84**, 739–749.

Preyer, S., Renz, S., Hemmert, W., Zenner, H. P., and Gummer, A. W. (**1996**). "Receptor potential of outer hair cells isolated from base to apex of the adult guinea-pig cochlea: Implications for cochlear tuning mechanism," Aud. Neurosci. **2**, 145–157.

Ratnanather, J. T., Spector, A. A., Popel, A. S., and Brownell, W. E. (**1996**). "Is the outer hair cell wall viscoelastic?," in *Proceedings of the Congress 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. 601–607.

Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.

Ruggero, M. A. (**1992**). "Responses to sound of the basilar membrane of the mammalian cochlea," Curr. Opin. Neurobiol. **2**, 449–456.

Russell, I. J., Cody, A. R., and Richardson, G. P. (**1986**). "The responses of inner and outer hair cells in the basal turn of the guinea-pig cochlea and in the mouse cochlea grown in vitro," Hear. Res. **22**, 199–216.

Russell, I. J., and Kossl, M. (**1992**). "Voltage responses to tones of outer hair cells in the basal turn of the guinea-pig cochlea: Significance for electromotility and desensitization," Proc. R. Soc. London, Ser. B **247**, 97–105.

Russell, I. J., and Nilsen, K. E. (**1997**). "The location of the cochlear amplifier: Spatial representation of a single tone on the guinea pig basilar membrane," Proc. Natl. Acad. Sci. U.S.A. **94**, 2660–2664.

Rybalchenko, V., and Santos-Sacchi, J. (**2003**). "Cl⁻ flux through a non-selective, stretch-sensitive conductance influences the outer hair cell motor of the guinea pig," J. Physiol. (London) **547**, 873–891.

Santos-Sacchi, J. (**1992**). "On the frequency limit and phase of outer hair cell motility: Effects of the membrane filter," J. Neurosci. **12**, 1906–1916.

Santos-Sacchi, J., Huang, G. J., and Wu, M. (**1997**). "Mapping the distribution of outer hair cell voltage-dependent conductances by electrical amputation," Biophys. J. **73**, 424–429.

Santos-Sacchi, J., Kakehata, S., Kikuchi, T., Katory, Y., and Takasaka, T. (**1998**). "Density of motility-related charge in the outer hair cell of the guinea pig is inversely related to best frequency," Neurosci. Lett. **256**, 155–158.

Scherer, M. M., and Gummer, A. W. (**2004**). "Impedance analysis of the organ of Corti with magnetically actuated probes," Biophys. J. **87**, 1378–1391.

Spector, A. A. (**2001**). "A nonlinear electroelastic model of the auditory outer hair cell," Int. J. Solids Struct. **38**, 2115–2129.

Spector, A. A., Ameen, M., Charalambides, P. G., and Popel, A. S. (**2002**). "Nanostructure, effective, properties, and deformation pattern of the cochlear outer hair cell cytoskeleton," J. Biomech. Eng. **124**, 180–187.

Spector, A. A., Brownell, W. E., and Popel, A. S. (**2003**). "Effect of outer hair cell piezoelectricity on high-frequency receptor potentials," J. Acoust. Soc. Am. **113**, 453–461.

Spector, A. A., Popel, A. S., Eatock, R. A., and Brownell, W. E. (**2005**). "Mechanosensitive channels in the lateral wall can enhance the cochlear outer hair cell frequency response," Ann. Biomed. Eng. **33**, 991–1002.

Tolomeo, J. A., and Steele, C. R. (**1995**). "Orthotropic piezoelectric properties of the cochlea outer hair cell wall," J. Acoust. Soc. Am. **97**, 3006–3011.

Tolomeo, J. A., and Steele, C. R. (**1998**). "A dynamic model of outer hair cell motility including intracellular and extracellular viscosity," J. Acoust. Soc. Am. **103**, 524–534.

Tolomeo, J. A., Steele, C. R., and Holley, M. C. (**1996**). "Mechanical properties of the lateral cortex of mammalian auditory outer hair cells," Biophys. J. **71**, 421–429.

Zheng, J., Shen, W., He, D. Z., Long, K. B., Madison, L. D., and Dallos, P. (**2000**). "Prestin is the motor protein of cochlear outer hair cell," Nature (London) **405**, 149–155.

Zwislocki, J. J., and Cefaratti, J. K. (**1989**). "Tectorial membrane. II. Stiffness measurements in vivo," Hear. Res. **42**, 211–227.

Modeling the cochlear nucleus: A site for monaural echo suppression?

Moritz Bürck[a] and J. Leo van Hemmen
Physik Department and BCCN—Munich, Technische Universität München, 85747 Garching bei München, Germany

Echo suppression plays an important role in identifying and localizing auditory objects. One can distinguish between binaural and monaural echo suppression, although the former is the one commonly referred to. Based on biological findings we introduce and analyze a mathematical model for a neural implementation of *monaural* echo suppression in the cochlear nucleus. The model's behavior has been verified by analytical calculations as well as by numerical simulations for several types of input signal. It shows that in the perception of a pair of clicks the leading click suppresses the lagging one and that suppression is maximal for an interclick interval of 2–3 ms. Similarly, ongoing stimuli will be affected by the suppression mechanism primarily a couple of milliseconds after onset, resulting in a reduced perception of a sound shortly after its start. Both effects match experimental data. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2770545]

I. INTRODUCTION

In any natural environment sound is reflected. The consequence is that every sound signal is followed by countless reflections, that is, echoes from many directions. This means that any biological system processing acoustic signals has to cope with reflections interfering with the original signal. Hence, a neuronal mechanism suppressing the confusing information of echoes is advantageous for sound perception.

Humans do not consciously perceive echoes arriving less than about 20 ms after the original signal.[1] A simple experiment shows that echo suppression is in part binaural and in part monaural. In a large room, for example a lecture room, one can perceive the otherwise suppressed echoes by covering one ear. In a small room, where the echoes are faster, this does not work.[2] Thus, neuronal echo suppression consists of a slower, binaural mechanism using both ears and a faster, monaural mechanism using only the cues one ear can provide. The focus of the research presented here is on the monaural mechanism for echo suppression.

II. MONAURAL ECHO SUPPRESSION: PSYCHOPHYSICAL AND BIOLOGICAL EVIDENCE

In 1963 Harris, Flanagan, and Watson investigated the binaural interaction of a click with a click pair.[3] In their experiment, a click pair was presented to one ear of the subjects while a single click was presented to their other ear. The subjects were asked to adjust the single click in time so that they perceive an auditory event straight ahead. The results imply that the second click of a click pair is not perceived, even unconsciously, if the interval between leading and lagging click is 2 ms. The second click *is* perceived if the interclick interval is *more* than 2 ms (in the experiment: 4, 6, and 8 ms) and, more interestingly, also if the interclick interval is

less than 2 ms (in the experiment: 0.5 and 1 ms). The authors conclude from these results that the suppression of the second click is therefore neither the result of a neuronal refractory period nor a function of mechanical properties of the basilar membrane. Instead they propose a *neural gate* that closes at about 1 ms after the start of the neuronal activity and opens after another 2 ms. Harris *et al.* suggested the cochlear nucleus as the site of this neural mechanism.

In 1980, Zurek referred to the idea of a neural gate as a possible explanation for the results he had obtained.[4] In his experiments, Zurek measured the time course of just-noticeable differences in interaural time and intensity using pairs of short (less than 5 ms) as well as long (50 ms) noise bursts. His principle finding was that the interaural sensitivity to changes in both time and intensity follows a nonmonotonic course after the onset of a sound. For about 0.5 to 10 ms after the beginning of the signal, the perception is degraded. *This degradation is maximal at* 1–5 ms.

A couple of years later, Hafter, in cooperation with several colleagues, published a number of experiments concerning the perception of clicks in a click series, especially as a function of click rates.[5–7] They found that clicks start influencing each other at interclick intervals shorter than 10 ms. This interference can be explained neither by neural refractoriness nor by narrow-band filtering within the auditory system nor by dependence of successive samples of internal noise. The authors suggested a "neural saturation process." This neural saturation process was the topic of further research where Hafter and colleagues have found the saturation to be a *monaural, frequency-specific* phenomenon.[7] As a location for the neural mechanism, they proposed the cochlear nucleus.

That is, the cochlear nucleus plays a key role in what we now call monaural echo suppression. Two major functional subdivisions of the cochlear nucleus complex are the anteroventral and the dorsal cochlear nucleus (AVCN and DCN). Both receive input from the auditory nerve (AN) and

---

a)Electronic mail: mbuerck@ph.tum.de
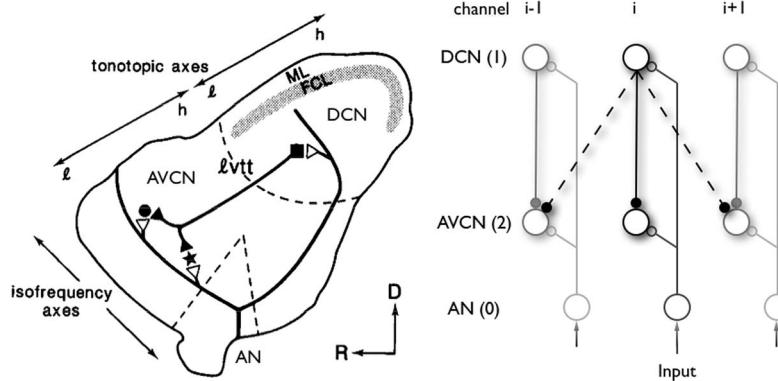
FIG. 1. Neuroanatomical structure in the cochlear nucleus as considered here. Left: sketch of Wickesberg and Oertel.[2] Both subdivisions of the cochlear nucleus are organized tonotopically. The auditory nerve (AN) is connected to AVCN (black circle and star) and DCN (black square) cells via excitatory synapses (white triangles). Cells of the DCN are connected to cells of the AVCN receiving input from the same fiber of the auditory nerve by inhibitory synapses (black triangles). Right: model of a neural frequency channel. Inhibitory synapses are represented by small black circles, excitatory synapses by small white circles. The inhibitory connections shown in dashed lines are considered in the numerical calculations only. They represent the spreading of inhibition into neighboring frequency channels, which is characterized by the quantity IS (*Inhibitory Spread*). For IS=1, there is no spreading, whereas for IS=n the neighboring $n-1$ nerve fibers receive inhibitory input. In the present figure (right), IS=3.

there is also a projection from the DCN to the AVCN. From the AVCN, the signals are transmitted to higher centers in the auditory brainsteam (see Fig. 1, left).

From 1988 to 1990, Wickesberg and Oertel carried out several elucidating experiments in the cochlear nucleus.[2,8] They investigated the properties of the projection from DCN to AVCN, which they found to be *frequency specific*. That means that cells in the DCN and in the AVCN that are connected to each other receive input from the same fiber of the auditory nerve. Furthermore, the projection was *inhibitory* for all cells in their experiments. The results are illustrated in Fig. 1, left. As for the timing of the inhibition, they found that inhibitory postsynaptic potentials reach the AVCN 2 ms after the stimulation of the auditory nerve. Excitatory postsynaptic inputs, however, reach the AVCN 0.7 ms after stimulation. The authors concluded that action potentials in the auditory nerve trigger an inhibition that can suppress later signals and that *the suppression is maximal if the delay of the second signal is* 2 ms. They proposed a role in echo suppression.

The remarkable match of biological and psychophysical findings has led us to model the circuitry Wickesberg and Oertel discovered, leading to a possible neural process of monaural echo suppression. The scope of the present work is thus to validate the intuitive interpretation of the biological evidence on a quantitative basis, providing an answer to the question of what the *intrinsic* properties of the neuronal projection from DCN to AVCN are.

III. MODELING THE COCHLEAR NUCLEUS

The layout of the model is simple (see Fig. 1, right). We have three different populations of neurons: the neurons of the auditory nerve (population 0), of the dorsal (pop. 1), and of the anteroventral cochlear nucleus (pop. 2). In order to keep the model as simple as possible, we will use identical parameters for the neurons in the AVCN and DCN for analytical as well as numerical calculations, in this way reducing the number of model parameters significantly. Our generic neurons can be considered to be bushy cells. This is not

far-fetched since Wickesberg and Oertel,[2] whom we owe our cell-parameter values, studied both bushy and stellate cells in the AVCN and found similar inhibitory response types among both cell types. The connections between the different populations have been modeled to match the findings of Wickesberg and Oertel[2,8] as well. Each neuron of the auditory nerve is connected excitatorily to one neuron of the DCN and to one neuron of the AVCN while the neuron in the AVCN is receiving inhibitory input from the cell in the DCN. Every three neurons that are connected as described above form one separate neuronal channel i. Such a channel is receiving input of only one frequency f. In other words, signal processing is frequency specific. Since a strict separation of frequencies may not be realized in actual biological systems, we allow the inhibition to spread into the neighboring channels (dashed lines in Fig. 1, right). Such a spreading of inhibition has indeed been reported in the DCN.[9,10] In order to find an analytical solution, however, we will assume strict frequency separation in the next subsection.

A. Analytical calculations

1. Methods

For our analytical calculations, we assume the firing of a neuron to be a statistical process. That is, an *inhomogeneous* Poisson process, which is defined by three properties. First, the probability of finding a spike between t and $t+\Delta t$ is $\lambda(t)\Delta t$, so $\lambda(t)$ is the time-dependent firing probability density or rate function. Second, the probability of finding two or more spikes there is $o(\Delta t)$, which means that we ignore their occurrence for small Δt. Third, a Poisson process has independent increments, i.e., events in disjoint intervals are independent.

A neuron i obeying such statistics with rate function

$$\lambda_i(t) = \nu^0 + \upsilon(t) = \nu^0 + \sum_{j,f} J_{ij}\epsilon(t - t_j^f - \Delta t_{ij}; \tau), \quad (1)$$

is then called a *Poisson neuron*; (see Refs. 11 and 12 for details). Here ν^0 is a spontaneous firing rate (which, for the sake of simplicity, we set to zero) and $\upsilon(t)$ is the cell poten-
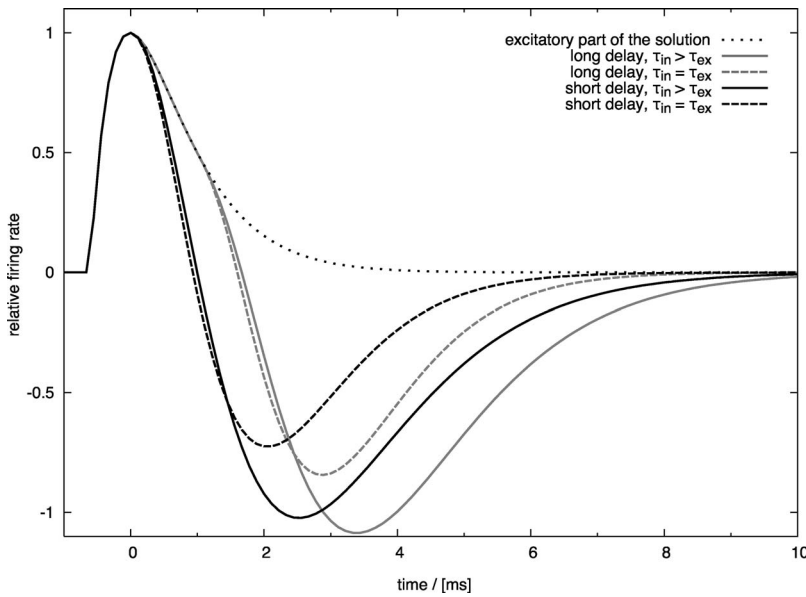
FIG. 2. Solution of Eq. (3) for a click (i.e., the impulse response). We have plotted the relative firing rate in arbitrary units versus time in milliseconds. The peaks of the solutions are aligned at 0 ms. Dotted line: excitatory part of the solution. Solid and dashed lines: solution for different time constants, black for a short and gray for a long delay of the inhibition. After rising, the firing rate drops below the spontaneous activity for all solutions, thus suppressing subsequent stimuli. The behavior is consistent with the psychophysical experiments[2] described before. $J_{10} = J_{20} = 1$, $J_{21} = -0.9$, $\Delta t_{20} = \Delta t_{10} = \Delta t_{21}$ $= 0.6$ ms (black lines) and 1.6 ms (gray lines) and τ_{ex} $= 0.6$ ms. $\tau_{in} = 0.6$ ms (dashed lines) and 1 ms (solid lines).

tial generating the spikes. It is determined by a sum over all input neurons j and their firing times t_j^f (abbreviated by f in the sum). The axonal delay between neuron j and neuron i is given by Δt_{ij}. Furthermore, J_{ij} is the synaptic strength of the coupling from neuron j to neuron i; it is positive for an excitatory and negative for an inhibitory coupling. The response of the cell potential to an incoming spike, the postsynaptic potential (PSP), is described by $\epsilon(t; \tau)$. We model the PSP for excitatory as well as inhibitory inputs by an alpha function,

$$\epsilon(t; \tau) = \begin{cases} (t/\tau)\exp(1 - t/\tau), & t \geqslant 0, \\ 0, & t < 0, \end{cases} \qquad (2)$$

with τ as the characteristic time constant, usually of the order of milliseconds. Note that τ may be different for inhibitory PSPs and excitatory PSPs (in our case $\tau_{ex} = 0.6$ ms or τ_{in} $= 1$ ms, respectively).

The firing rate as defined above refers to a stochastic quantity. We are interested in the expectation value of the rate rather than a specific realization of the stochastic firing process since the process considered here is self-averaging.[12] It can be shown that instead of summing over the spikes in Eq. (1), we can then integrate over the expectation value of the input spikes.[12] As a consequence, the expectation value of the rate function of the neurons in the AVCN is

$$\lambda_2(t) = J_{20} \int_0^\infty ds\, \epsilon(s; \tau_{ex}) F(t - \Delta t_{20} - s)$$

$$+ J_{10}J_{21} \int_0^\infty ds\, \epsilon(s; \tau_{ex}) \int_0^\infty ds'(s'; \tau_{in})$$

$$\times F(t - \Delta t_{10} - \Delta t_{21} - s - s'), \qquad (3)$$

where we assume that the rate function of the neurons in the auditory nerve is proportional to the input function $F(t)$. We define an input function to be the firing rate in the auditory nerve induced by an external stimulus.

In the case of identical time constants for excitatory and inhibitory PSPs ($\tau := \tau_{ex} = \tau_{in}$), the resulting "impulse response" (the solution to a delta function as approximation of a click) is given by

$$\lambda_2(t) = J_{20}\epsilon_\tau(t - \Delta t_{20}) + J_{10}J_{21}(\epsilon_\tau * \epsilon_\tau)(t - \Delta t_{10} - \Delta t_{21}),$$

$$(4)$$

with $\epsilon_\tau(t) := \epsilon(t; \tau)$ and "*" denoting the convolution; calculated explicitly,

$$\lambda_2(t) = J_{20}\epsilon_\tau(t - \Delta t_{20}) + J_{10}J_{21} \frac{e^1(t - \Delta t_{10} - \Delta t_{21})^2}{6\tau}$$

$$\times \epsilon_\tau(t - \Delta t_{10} - \Delta t_{21}). \qquad (5)$$

The first term on the right in (3)–(5) represents the excitatory influence of auditory-nerve activity and the second term stands for the inhibitory influence of the activity in the DCN, which is driven itself by an excitatory input from the auditory nerve. In (4) and (5) we can see that the inhibition, in comparison to excitation, is delayed by the time shifting and smeared out by the convolution with the second PSP. It nevertheless follows a similar time course as the excitation, viz., that of an EPSP. The dashed lines in Fig. 2 show realizations of (4) and (5) (see the next section for a more detailed description).

We have solved (3) for two input functions, one representing a click and one representing the onset of a pure tone. We have chosen the delta function to simulate the click, viz., $F(t) = \delta(t)$. As for the pure tone, we had to choose a slightly more complex function. The onset is described by a Heaviside stepfunction; $\Theta(t) = 1$ for $t \geqslant 0$ and 0 for $t < 0$. For an oscillation we have to take into account that only the compressional part of the signal, that means an increase in the air pressure, evokes neural activity in the auditory system which, in addition, adapts rapidly to an ongoing signal. So, the half-wave rectified part of a sine is used to model the activity in the auditory nerve in response to a pure tone in combination with a decaying exponential term. Since this would lead to quite a bulky expression in the analytical so-

M. Buerck and J. van Hemmen: Modeling monaural echo suppression
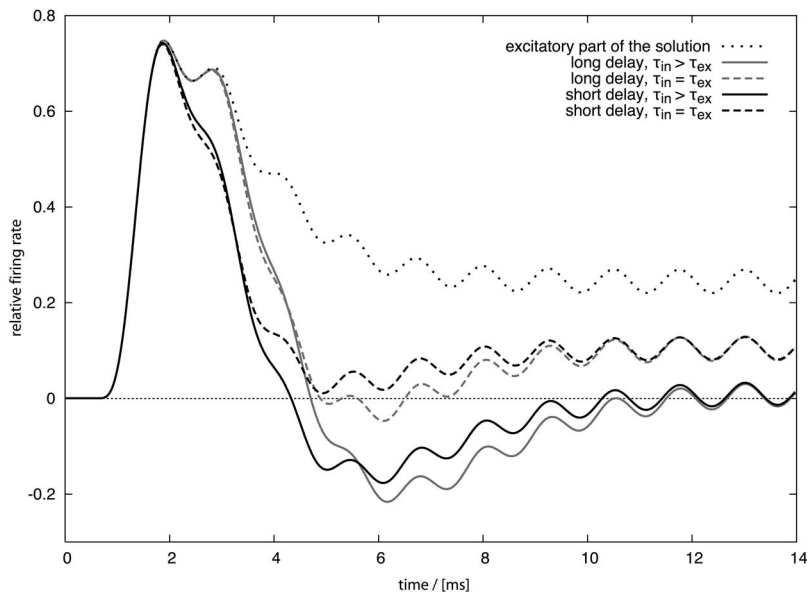
FIG. 3. Solution of Eq. (3) for the onset of a pure tone. We plot the relative firing rate in arbitrary units versus time in milliseconds. The curves are aligned similarly to Fig. 2. Dotted line: excitatory part of the solution. Solid and dashed lines: solution for different time constants, black for a short and gray for a long delay of the inhibition. The inhibition results in a drop of the firing rate shortly after the onset of the signal, which is equivalent to the decreased sensitivity after the beginning of a signal reported in psychophysical experiments.[4] Note that the amplitude of the oscillation is identical for all three curves, which means that the phase information of the signal is not influenced by the inhibition. $J_{10}=J_{20}=1$, $J_{21}=-0.35$, $\Delta t_{20}=\Delta t_{10}=\Delta t_{21}=0.6$ ms (black lines) and 1.6 ms (gray lines), $\tau_{ex}=0.6$ ms and $f=800$ Hz. $\tau_{in}=0.6$ ms (dashed lines) and 1 ms (solid lines).

lution, yet another function was used, $F(t)=[1-\cos(\omega t)]\,[0.15+0.85\exp(-t/\tau_{adapt})]$ with $\tau_{adapt}=1.1$ ms.[30] By using this function we assume the neurons in the auditory nerve to be slower and less precise than they actually are.

2. Results

Four solutions for a delta-function click as input are shown in Fig. 2 in comparison to their purely excitatory part (solid and dashed lines in comparison to the dotted line). The solid lines represent the solutions for identical time constants of inhibition and excitation whereas for the dashed lines the time constant of the inhibition was larger. Based on Wickesberg and Oertel's *in vitro* experiments,[2] the axonal delay from DCN to AVCN was set to 1.6 ms (gray lines). The shorter axonal delay of 0.6 ms (black lines) is in accordance with *in vivo* experiments performed by Wickesberg in 1996.[13]

After the click has been fed into the neural network, the firing rate begins to increase slowly, then grows faster and reaches a clear maximum before decreasing again. For the relevant case, viz., the case of a short axonal delay (0.6 ms), the inhibition results in a significant sharpening of the peak as discovered before experimentally[13] and in a drop of the firing rate below the spontaneous firing rate shortly after the signal. Most important, there is a *pronounced minimum* of the firing rate corresponding to a maximum of the inhibition. For identical time constants of excitation and inhibition, this minimum is found at about 2 ms after the firing rate reaches its maximum (dashed black line in Fig. 2). As the time constant of the inhibition increases, the minimum moves away from the maximum (solid black line in Fig. 2) and gets stronger. We note that a longer axonal delay from auditory nerve to AVCN via DCN ($\Delta t_{10}+\Delta t_{21}$) does not influence the strength or the form of the excitation or inhibition, but only shifts the start of the IPSP in time (gray lines in Fig. 2).

It is important to realize that except for the weights, which have not been determined physiologically, the parameters used to obtain the results are not the result of any arbitrary fitting. They were, instead, taken from the publications

by Wickesberg and Oertel[2] and by Wickesberg[13] and are in the typical range for neurons in the cochlear nucleus.[14–16]

Since the system is effectively linear, the solution for two subsequent clicks is the sum of the solutions for the individual clicks. That means *a second click following the first one will be suppressed* and the suppression will be maximal at a delay of 2 to 2.5 ms, depending on the time constant of the inhibition. The behavior of the model thus matches the experimental results of Harris, Flanagan, and Watson very well.

As for a pure tone, several solutions for a beginning tone of 800 Hz are shown in Fig. 3. All the solutions have in common the drop of the firing rate shortly after the onset of the signal due to the inhibition. For the solutions with a large inhibitory time constant, the firing rate drops below the spontaneous firing rate, consistent with psychophysical data indicating decreased sensitivity immediately after the onset of an ongoing stimulus.[4] Then the firing rate increases and reaches a level that depends on the time constant of the inhibition. By a change of the strength of the inhibition, the time constant therefore shifts the overall level of activity, which is an effect of the normalization of the PSPs. On closer examination, however, it becomes evident that the amplitude of the oscillation is identical for all three graphs plotted in the figure. This means that the phase information of the signal is preserved. Thus the inhibition does not degrade the information on the frequency of the pure tone but, in our model, even improves the signal-to-noise ratio.

Altogether, the behavior of the model matches the psychophysical results cited above. A single click will suppress a subsequent click and the suppression is maximal a couple of milliseconds after the first click.[3] The perception of an ongoing stimulus is distorted shortly after its onset, whereas its perception later on is not disturbed.[4]

B. Numerical calculations

The consistency of our solution and the psychophysical evidence is remarkable, but it is still hard to estimate how close our analytical results are to the behavior of the real

biological system. Though the parameters used are biologically plausible, our neuron model has ignored an essential characteristic of any neuron, the threshold. The threshold, however, will play an important role at the beginning of any signal processed so that our results are encouraging but do not necessarily provide a biologically relevant explanation of the effect described in Sec. I. In the next section, we will use another method for modeling the neurons in our system. The new model will feature an explicit threshold and therefore overcome the limitations of the Poisson neuron. The trade-off is in the loss of the analytical solution.

1. Methods

The model we use in this section is an extension of the leaky integrate-and-fire neuron, the *spike response model* (SRM).[12] In the SRM description, spikes are represented by delta functions and absolute as well as relative refractory periods are taken explicitly into account. The cell potential $v_i(t)$ of a cell i receiving input from the cell j is given by

$$v_i(t) = \sum_f \eta(t - t_i^f) + \sum_{j,f} J_{ij}\epsilon(t - t_j^f - \Delta_{ij}^{\text{axon}}). \quad (6)$$

As before, the sum over f is a sum over the firing times t^f of the neurons. The new terms $\eta(t - t_i^f)$ represent the refractory behavior of the neuron i we are focussing on and are described by

$$\eta(t) = \begin{cases} -\infty, & t < \tau_{\text{abs}}, \\ -N \exp(-t/\tau_{\text{rel}}), & t \geq \tau_{\text{abs}}. \end{cases} \quad (7)$$

In our simulations τ_{abs} is taken to be 0.25 ms and N equals two times the maximum value a single EPSP reaches. As before, J_{ij} is the strength of the coupling from neuron j to neuron i and ϵ is the postsynaptic response to a spike. Again τ will be different for excitatory and inhibitory PSPs (0.6 and 1 ms). The axonal delay $\Delta_{ij}^{\text{axon}}$ is composed of the synaptic delay and the finite speed of propagation of a spike along the axon. A spike will be fired by neuron i if the potential $v_i(t)$ exceeds a threshold ϑ which is set to 0.9 of the maximum value an EPSP reaches.

Consequently, for both parameter sets the DCN is primarily acting as a relay of the AN spikes, delaying them temporally and making their action inhibitory. The free parameters are the strength of the inhibition, J_{in}, and the spreading of the inhibition in neighboring frequencies, IS (*I*nhibitory *S*pread). For IS=1, there is no spreading whereas for IS=n, the neighboring $n-1$ nerve fibers will receive inhibitory input (see Fig. 1). The strength of the inhibition decreases exponentially with the distance between the nerve fibers. If not specified otherwise, we have used a minimal spreading (IS=5) and a medium strength of inhibition (J_{in}=−0.8).

The cell parameters were extracted from Refs. 2 and 17. With regard to the different values for the relative delay of inhibition with respect to excitation in the AVCN found *in vitro*[2] (1.6 ms) and *in vivo*[13] (0.6 ms), we have checked the behavior of our model for both short and long delay (data not shown). In addition we have tested the robustness of the

TABLE I. The parameters as used for the numerical simulations. The first parameter set is consistent with Wickesberg and Oertel[2], Wu and Oertel[17], and Wickesberg[13]. The parameters in brackets lead to very similar (axonal delay) or virtually identical behavior (time constant of the relative refractory period) as the primary parameter set (data not shown). The strength of the inhibition was varied to optimize the behavior of the model.

τ_{rel}	0.3(1.0) ms
τ_{ex}	0.6 ms
τ_{in}	1 ms
τ_{abs}	0.25 ms
ϑ	0.9
J_{ex}	1
J_{in}	−0.8
$\Delta_{ij}^{\text{axon}}$	0.6(1.6) ms

model through two different values for the time constant of the relative refractory period, τ_{rel}=0.3 ms as it has been measured by Wu and Oertel[17] and τ_{rel}=1 ms (data not shown). The parameters are summarized in Table I.

As for modeling the periphery of the auditory system, we have taken advantage of the C++ package LUTEar,[18] which has been developed at the university of Essex. LUTEar reproduces the spike times in the auditory nerve fibers corresponding to any wav-file. Version 2.0.9 was used to calculate the spike times in 500 frequency channels, which cover a frequency range from 200 Hz to 16 kHz. The channels are distributed uniformly on a logarithmic scale, resulting in a doubling of frequency every 79 channels. Each frequency channel consists of one nerve fiber. The output of one nerve fiber is the input for one cell in the AVCN and DCN of our model (Fig. 1). We use the spike times of clicks, pairs of clicks, and white noise as well as a pure tone (data not shown).

2. Results

As in the analytical calculations, the behavior of the model is consistent with the results of Harris, Flanagan, and Watson.[3] Figures 4 and 5 illustrate the behavior of the model for a click series. A single click is followed by eight pairs of clicks with interclick intervals (ICIs) of 0.5, 1, 2, 3, 4, 6, 8, and 10 ms, each pair being separated from its neighboring pairs by 40 ms.

As the interclick interval increases in the auditory nerve (Fig. 4), the bands of activity become broader until, at a separation of 2 ms, we can distinguish the individual clicks. Every click is followed by an oscillating activity in the lower frequencies, which is the result of natural oscillations of the basilar membrane. At the stage of the cochlear nucleus, these oscillations almost disappear (see Fig. 5).

Furthermore, we notice that the overall activity is reduced almost by a factor of 2 in comparison to the activity in the auditory nerve [15 794 spikes in the AN (Fig. 4) and 8893 in the AVCN (Fig. 5)]. At a second glance, it becomes apparent that the first four pairs of clicks are indistinguishable. It is only at an interclick interval of 4 ms that the second click appears, which gets stronger as the separation between the clicks increases. In Figs. 4 and 5, the time constant of the relative refractory period is 0.3 ms and the axonal delay is 0.6 ms, but for τ_{rel}=1 ms the graphical representa-

M. Buerck and J. van Hemmen: Modeling monaural echo suppression
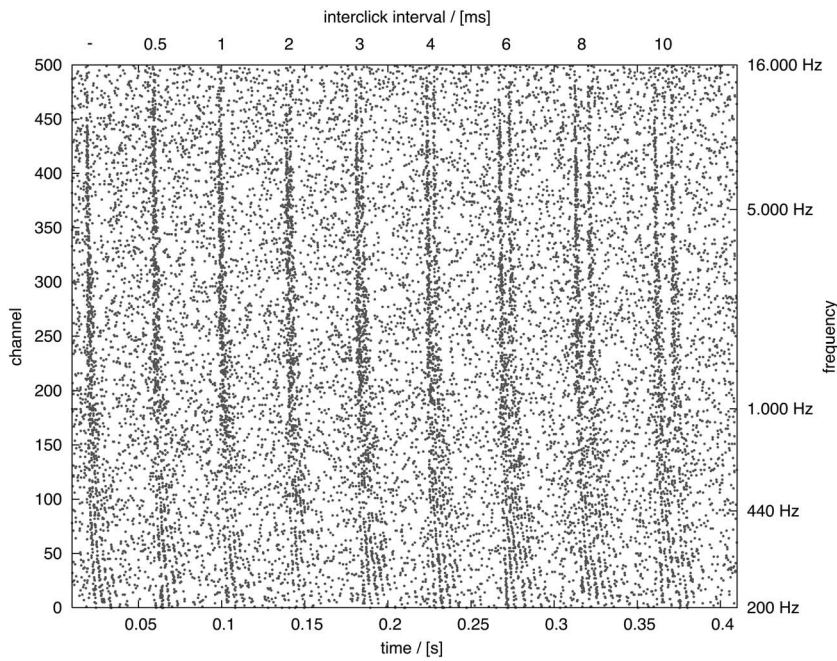
FIG. 4. Spike times for a series of clicks in the auditory nerve. The horizontal axis denotes time, whereas the vertical axis exhibits channel number (left) and corresponding frequency (right). One single click is followed by eight pairs of clicks with an interclick interval of 0.5, 1, 2, 3, 4, 6, 8, and 10 ms. Each pair of clicks is separated from the neighboring pairs by 40 ms of silence. The activity corresponding to the individual clicks as well as the natural oscillation of the basilar membrane can be discerned clearly. Calculated with LUTEar.[18]

tion of the neural activity in the cochlear nucleus is practically indistinguishable (data not shown). Increasing the axonal delay to 1.6 ms leads to a very similar result where the second click is suppressed for ICIs of 2 and 3 ms but at the same time the response to the single click as well as to the click pairs for an ICI of 0,5 and 1 ms is slightly broader in time (data not shown). An increase in the spreading of the inhibition (IS > 5) results in an almost identical behavior if the strength of the inhibition is reduced simultaneously (data not shown).

The spike histogram in Fig. 6 provides a more detailed view of the click suppression illustrated in Figs. 4 and 5. We see the spike activity evoked by a single click (light transparent gray) compared to that evoked by a click pair (dark gray) for the same set of interclick intervals as used by Harris *et al.*[3] in the auditory nerve (left) and the cochlear nucleus

(right). The data have been obtained by adding ten runs and calculating the running average over 0.0228 ms in the upper 250 frequency channels, covering the frequency range from 1790 Hz to 16 kHz. In the auditory nerve the two clicks of a click pair can be discerned clearly for all interclick intervals but do not form distinct events for ICIs of 1 and 0.5 ms. In the cochlear nucleus the second click is suppressed for interclick intervals of 2 and to some extent also of 1 ms but not of 0.5 ms (width at half height is 1.7 ms for an ICI of 2 ms, 1.8 ms for an ICI of 1 ms, and 1.7 ms for a single click). The suppression in our model is therefore not due to the refractory period of the neurons and our results agree with the results of Harris *et al.*[3] in that the second click of a click pair can be perceived for *very* short and medial but not for short interclick intervals. We note that Harris *et al.*[3] reported that the second click of a click pair is perceived consciously only
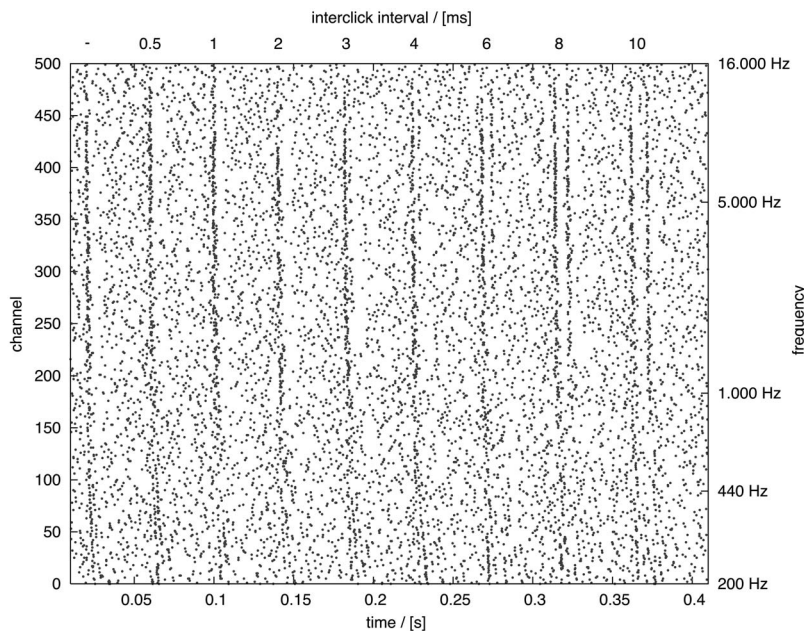
FIG. 5. Spike times for a series of clicks in the cochlear nucleus, plotted in a way identical to the one of Fig. 4. The natural oscillations of the basilar membrane almost vanish and the first three pairs of clicks are indistinguishable from each other. Especially the click pair with an interclick interval of 2 ms differs from that of Fig. 4 in that the second click has disappeared clearly. The behavior of the model matches the results of Harris *et al.*,[3] where the second click is not perceived at an interclick interval of 2 ms. Δt_{12}=0.6 ms, τ_{rel}=0.3 ms, and IS=5. The inhibitory spread (IS) has been defined in Fig. 1.
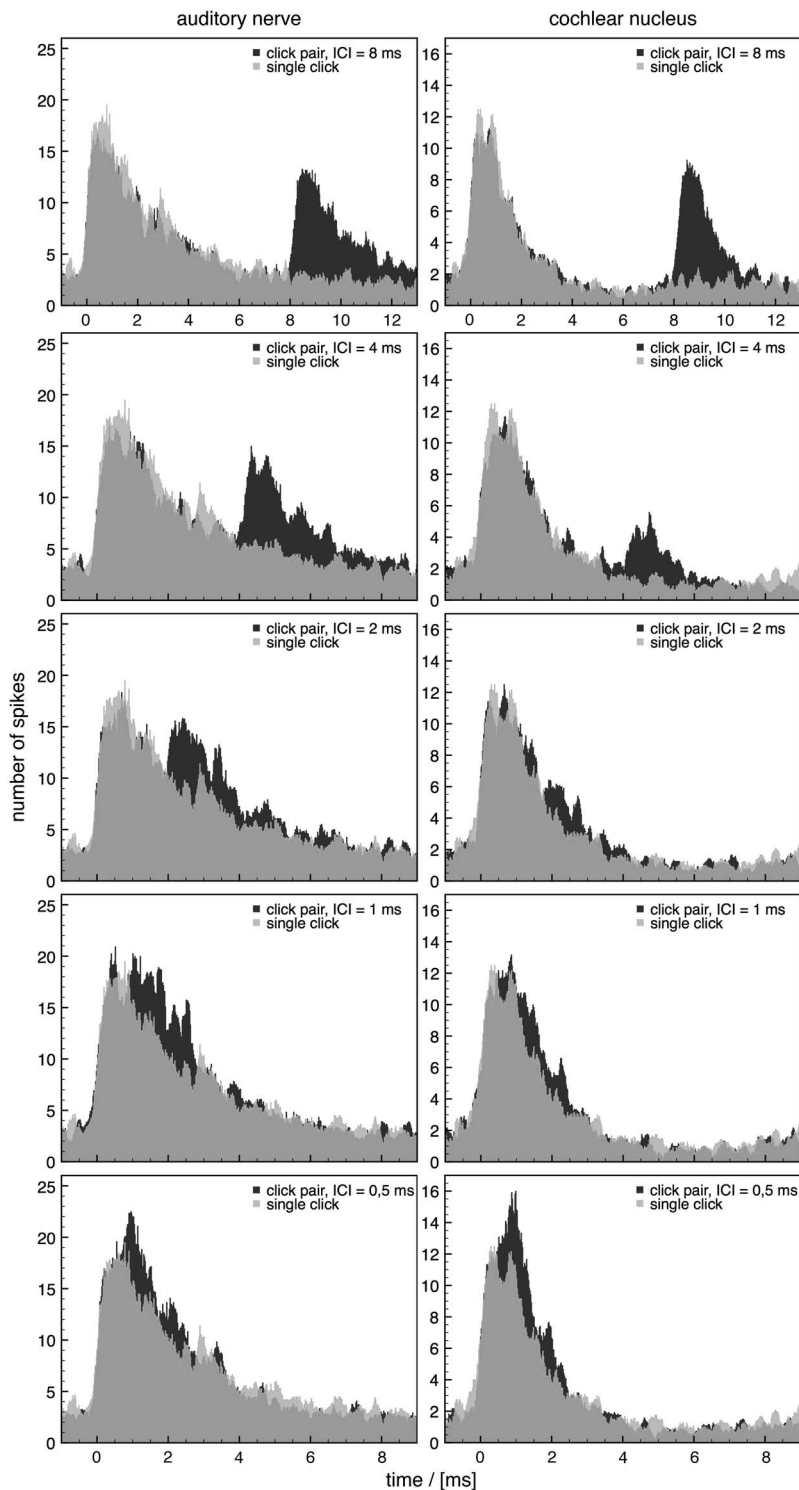
FIG. 6. Comparison of the spike activity evoked by a single click (light transparent gray) to that evoked by click pairs (dark gray) in the auditory nerve (left) and in the cochlear nucleus (right) for interclick intervals (ICI) of 0.5, 1, 2, 4, and 8 ms (bottom to top). The data have been obtained by adding several runs (ten repetitions, bin size=0.0019 ms) and calculating a running average (12 bins=0.0228 ms); only the activity in channels 250 to 500 (1790 Hz to 16 kHz) has been taken into account. All pairs of clicks can be discerned clearly from a single click in the auditory nerve, whereas in the cochlear nucleus the second click of a click pair is suppressed at an ICI of 2 ms. The visible presence of the second click at an ICI of 1 and 0.5 ms in the cochlear nucleus proves that the suppression at an ICI of 2 ms is not due to a neuronal refractory period or the like. Parameter values are $\Delta t_{12}=0.6$ ms, $\tau_{\mathrm{rel}}=0.3$ ms, and IS=5.

at interspike intervals of 4 and 8 ms and that in our simulations the two clicks form a clearly distinct event only at the very same click delays.

A closer examination of the activity corresponding to a single click reveals two peaks of activity in the cochlear nucleus but *not* in the auditory nerve (e.g., top row of Fig. 6). Interestingly, 1 ms after a *single click*, that is even when there was *no second click*, a click was perceived in a significant portion of the trials in Harris *et al.*'s experiment.[3] The suppression of activity in the trial of each click results in a

sharpening of the peaks, which was described by Wickesberg before.[13]

In addition to the response of the model to a click series, we have evaluated the response to a pure tone (data not shown) and to white noise. In order to understand the behavior of the model we have analyzed at the interspike interval distribution in the cochlear nucleus across all frequency channels during half a second amidst a lasting pure tone of 440 Hz (data not shown) and ongoing white noise. The sound pressure level was 100 dB(A), which is relatively loud
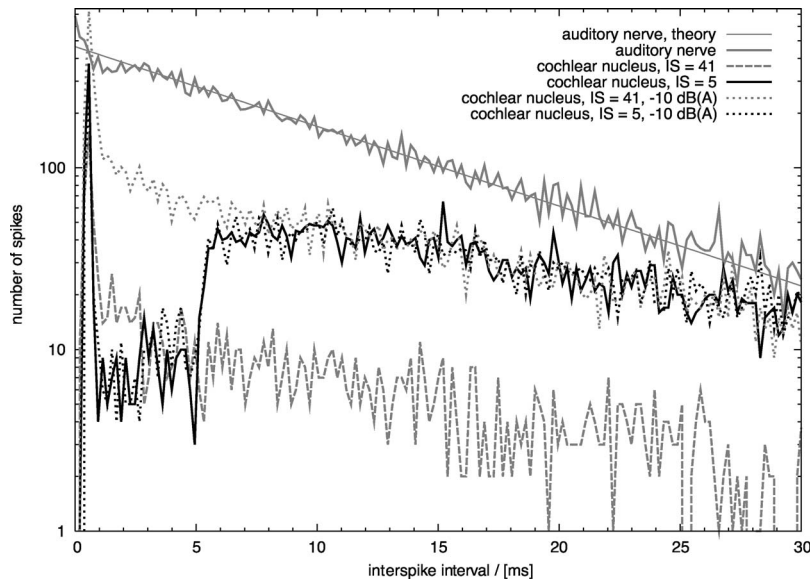
FIG. 7. Response to white noise. Distribution of spikes summed over all frequency channels in the auditory nerve and the cochlear nucleus, plotted as a response (vertical axis) to ongoing white noise in dependence upon interspike intervals (horizontal axis). Solid gray lines: activity in the auditory nerve (thick line) and the theoretical interspike interval distribution for random spikes (straight thin line). Dashed gray and solid black line: activity in the cochlear nucleus. Dotted gray and black lines: activity in the cochlear nucleus for a sound pressure level reduced by 10 dB (A). In the case of small spreading of the inhibition (IS) into neighboring frequency channels, spikes with interspike intervals less than 5 ms are subject to significant inhibition whereas spikes with larger interspike intervals are only slightly inhibited (black line). This is consistent with psychophysical data indicating that the perception of an ongoing stimulus is degraded a few milliseconds after its onset.[4] Greater spreading inhibition results in equal spike suppression for both small and large interspike intervals; the characteristic inhibition at small interspike intervals is lost (dashed and dotted gray line). A reduction in the sound pressure level results in a significant increase of activity for larger inhibitory spreading (dashed and dotted gray line) but does not change the activity for small inhibitory spreading (solid and dotted black line). Parameter values are $\tau_{rel}=0.3$ ms, $J_{in}=-0.8$, and IS=5 (black lines) and $J_{in}=-0.05$ and IS=41 (gray dashed and dotted line). IS=41 means that the inhibition spreads into the neighboring 40 frequency channels where 79 channels correspond to one octave (see also Fig. 1).

but improves the statistics by increasing the number of spikes. The interspike interval distribution in Fig. 7 represents ongoing white noise; the distribution for a lasting pure tone bears the same characteristics. The gray solid lines represent the interspike interval distribution in the auditory nerve, the thick line the actual data and the thin line the distribution expected for a stochastic point process. The black solid line corresponds to the distribution in the cochlear nucleus with IS=5 and the dashed gray line to that with IS=41. The dotted lines represent the spike activity in response to white noise at a sound pressure level of 90 dB(A), the gray dotted line for a large (IS=41) and the black dotted line (almost identical to the solid black line) for a small (IS=5) inhibitory spreading.

The large inhibitory spread of about half an octave (IS=41) leads to a decreased overall activity, a reduction of the signal intensity similar to the analytical results obtained before. Compared to the activity in the auditory nerve, the characteristics of the distribution change only slightly. This contrasts with the result for a small spreading of the inhibition into neighboring frequencies (IS=5). Here the overall intensity is reduced as well, but spikes with interspike intervals less than 5 ms are suppressed disproportionately to spikes with larger interspike intervals. The drop in activity at small interspike intervals corresponds well with reduced sensitivity in the perception of an ongoing stimulus shortly after its onset.[4] For a couple of milliseconds after onset, there are simply no spikes left to evaluate. Note the peak at an interspike interval of about 0.5 ms, which agrees with Fig. 6. A rapid succession of spikes can evoke more spikes than the

succession contains because the time constant of the relative refractory period is much smaller than the time constant of the EPSP. Accordingly, the peak at 0.5 ms shrinks for a longer time constant of the relative refractory period (data not shown). Interestingly, analogous simulations with a lasting pure tone show that the synchronization of the neuronal activity to the signal changes only slightly, the vector strength being 0.71 in the AN and 0.74 in the AVCN (data not shown). When the sound pressure level of the signal is reduced by 10 dB(A) the number of spikes remains constant in the case of a small inhibitory spread (compare dotted and solid black line in Fig. 7) but increases in the case of a large inhibitory spread (dotted and dashed gray line in Fig. 7). The model thus provides nonlinear gain control for frequency-specific inhibition.

We note that the equivalence of a broad weak and a narrow strong inhibition as we found it for click pairs breaks down for a narrow-band signal (as a pure tone) or a random signal (as white noise).

IV. DISCUSSION

In the previous two sections the analytical as well as the numerical approach have led to the same conclusions concerning the properties of the model presented. In both cases the second click of a click pair is suppressed and the suppression is maximal at about 2 ms after the first click (Figs. 2 and 6). As for the response to the onset of an ongoing stimulus, the spiking activity is minimal a couple of milliseconds after onset (Figs. 3 and 7). This behavior is consis-

tent with the psychophysical experiments cited in Sec. I. Moreover, it is very robust. The only free parameters are the strength of the inhibition and the firing threshold of the neurons. All other parameters arise naturally from matching both form and amplitude of the model's cell potentials to the measurements published by Wickesberg and Oertel,[2] Wickesberg,[13] and Wu and Oertel.[17] Hence our model is highly plausible from a biological point of view.

Whereas the model can reproduce the psychophysical results obtained for clicks using both a relatively strong, frequency-specific inhibition and a weaker inhibition spreading across many frequency channels, a strong, frequency-specific inhibition is needed in order to reconstruct psychophysical results for ongoing stimuli. This can be understood by the following argument: To suppress a single spike in one frequency channel, a certain amount of inhibition is needed. This inhibition can arise either from a strong frequency-specific inhibitory connection (small IS) or from an integration over many weak inhibitory inputs (large IS). In the case of a precisely timed broadband signal as a click both are possible, but as either broadband character (in case of a pure tone) or precise timing (in case of white noise) disappear, a large inhibitory spread will yield only nonspecific spike suppression.

As for the function of the frequency-specific inhibitory connection from the DCN to the AVCN, our results suggest three possibilities.

First, analytical (Fig. 3) and numerical (Figs. 4, 5, and 7) calculations show that a type of *gain control* is provided since the response is reduced significantly through inhibition. The model can even provide constant output for different input intensities in the case of an ongoing stimulus.

Second, the analytical calculations show that inhibition induces *contrast enhancement* since the dc component of the pure tone response is decreased, whereas the ac component (i.e., the amplitude of the oscillation) is unchanged (Fig. 3). This means that phase locking should be improved since the ac to dc ratio is larger with inhibition than without. An improvement in phase locking has been reported physiologically in the AVCN as compared to the AN.[19] In our numerical simulations this effect was only marginal. Presumably this is due to the fact that multiple AN fibers need to converge onto an AVCN cell in order to show enhanced synchronization, as Joris *et al.* demonstrated in their coincidence cell model.[19] Our model, however, only specifies a single AN fiber innervating a single AVCN cell.

Third, *monaural echo suppression* could be an additional function of the inhibitory circuitry we have modeled in the cochlear nucleus. A click quickly following another click will under natural circumstances most often be an echo. It is known that echoes occurring 2–20 ms after the initial signal are suppressed in the central nervous system.[1] A model for *binaural* echo suppression has been proposed previously in this journal[20] and has been extended in a dissertation available online.[21] Both versions of the model feature a persistent inhibition that suppresses echoes for up to 20 ms. A drawback of the extended version of the binaural model is a minimal suppression time of 5 ms, which means it is not capable of a suppression occurring in less than 5 ms. That is, our model with a suppression time of about 2–4 ms completes the picture.

Apart from the functions listed above, several publications hint towards yet an additional role delayed, frequency-specific inhibition might play. A neuronal circuitry very similar to the one discussed in the present paper has been found in the medial superior olive of the mustached bat.[22] There a delayed monaural inhibition is proposed as a kind of filter for amplitude-modulated (AM) sounds, important to detecting and identifying wing-beating prey. Amplitude modulations are known to play an important role in the processing of many acoustic signals. They are of special importance to transmitting information in speech[23] as well as for the perception of pitch.[24] It has been shown that identifying acoustic signal periodicity is well within the scope of neuronal hardware.[25]

An idea related to a delayed monaural frequency-specific inhibition was brought up by de Cheveigné in 1998[26] in the context of pitch perception. He showed that many pitch phenomena can be explained by a model involving an array of delay lines and inhibitory gating. Against this background, it is very interesting that in 1996 Wickesberg found a second, even faster mechanism of inhibition in the cochlear nucleus.[13] Referring to observations of Shore,[27] Wickesberg speculated that "the circuitry in the cochlear nucleus may be just the first stage in an ascending and descending[27] cascade of delayed, on-frequency inhibition." It is an important fact that the circuitry in the real cochlear nucleus does not only extend to the bushy cells that project to the superior olive complex and are involved in sound localization (where echo suppression is essential), but also to the stellate cells that project to the inferior colliculus. Stellate cells have much slower responses than bushy cells,[14] which means inhibition would follow excitation more closely. Inhibiting different types of cells could contribute to extending the range of the timing in a cascade of delayed inhibition, so we could think of a cascade of frequency-specific inhibitory loops filtering signals according to their respective AM frequency. Indeed, a phenomenological model of responses to AM tones uses the circuitry we have analyzed in the present paper as a module.[28]

Finally, it should be mentioned that the filtering of a signal obviously performed by the cochlear nucleus shows remarkable similarities with a filtering technology used in sound studios and extensively in radio as well as TV broadcasts.[29] The so-called *compressors* attenuate the signal after it reaches a certain threshold. That is, they reduce the dynamic range of the signal. If such a compressor is set, for example, to an "attack" of 2 ms and a "decay" of 10 ms, it will start to attenuate the signal 2 ms after the amplitude of the audio signal exceeds its threshold. This attenuation will gradually decay back to 0 within the next 10 ms, a behavior similar to that of our model. A compressor improves the perceived volume of an audio signal.

We conclude that the neural circuitry discovered by Wickesberg and Oertel[2] in the cochlear nucleus—in addition to providing gain control—can be regarded as the site of monaural echo suppression. Furthermore, it may be involved in the preprocessing of amplitude-modulated sounds.

[1] J. Blauert, *Spatial Hearing* (MIT, Cambridge, MA, 1999).

[2] R. E. Wickesberg and D. Oertel, "Delayed, frequency-specific inhibition in the cochlear nuclei of mice: A mechanism for monaural echo suppression," J. Neurosci. **10**, 1762–1768 (1990).

[3] G. G. Harris, J. L. Flanagan, and B. J. Watson, "Binaural interaction of a click with a click pair," J. Acoust. Soc. Am. **35**, 672–678 (1963).

[4] P. M. Zurek, "The precedence effect and its possible role in the avoidance of interaural ambiguities," J. Acoust. Soc. Am. **67**, 952–964 (1980).

[5] E. R. Hafter and R. H. Dye, "Detection of interaural differences of time in trains of high frequency clicks as a function of interclick interval and number," J. Acoust. Soc. Am. **73**, 644–651 (1983).

[6] E. R. Hafter and E. M. Wenzel, "Lateralization of transients presented at high rates: site of the saturation effect," in *Hearing—Physiological Basis and Psychophysics*, edited by R. Klinke and R. Hartman (Springer, Berlin, 1983), pp. 208–220.

[7] E. R. Hafter, T. N. Buell, and V. M. Richards, "Onset coding in lateralization: Its form, site and function," in *Auditory function*, edited by G. Edelman (Wiley, New York, 1988), pp. 647–674.

[8] R. E. Wickesberg and D. Oertel, "Tonotopic projection from the dorsal to the anteroventral cochlear nucleus of mice," J. Comp. Neurol. **268**, 389–399 (1988).

[9] G. A. Spirou, K. A. Davis, I. Nelken, and E. D. Young, "Spectral Integration by Type II Interneurons in Dorsal Cochlear Nucleus," J. Neurophysiol. **82**, 648–663 (1999).

[10] G. A. Spirou and E. D. Young, "Organization of dorsal cochlear nucleus type IV unit response maps and their relationship to activation by band-limited noise," J. Neurophysiol. **66**, 1750–1768 (1991).

[11] R. Kempter, W. Gerstner, and J. L. van Hemmen, "Hebbian learning and spiking neurons," Phys. Rev. E **59**, 4498–4514 (1999).

[12] J. L. van Hemmen, "Theory of synaptic plasticity," in *Handbook of Biological Physics (Vol. 4), Neuro-informatics, Neural Modelling*, edited by F. Moss and S. Gielen (Elsevier, Amsterdam, 2001), pp. 771–823.

[13] R. E. Wickesberg, "Rapid inhibition in the cochlear nuclear complex of the chinchilla," J. Acoust. Soc. Am. **100**, 1691–1702 (1996).

[14] J. S. Isaacson and B. Walmsley, "Receptors underlying excitatory synaptic transmission in slices of the rat anteroventral cochlear nucleus," J. Neurophysiol. **73**, 964–973 (1995).

[15] A. J. Smith, S. Owens, and I. D. Forsythe, "Characterisation of inhibitory and excitatory postsynaptic currents of the rat medial superior olive," J. Physiol. (London) **529**, 681–698 (2000).

[16] M. Barnes-Davies and I. Forsythe, "Pre- and postsynaptic glutamate receptors at a giant excitatory synapse in rat auditory brainstem slices," J. Physiol. (London) **488**, 387–406 (1995).

[17] S. H. Wu and D. Oertel, "Intracellular Injection with Horseradish Peroxidase of Physiologically Characterized Stellate and Bushy Cells in Slices of Mouse Anteroventral Cochlear Nucleus," J. Neurosci. **4**, 1577–1588 (1984).

[18] L. P. O'Mard, M. J. Hewitt, and R. Meddis, "Lutear, v2.0.9," C++ core routines library, University of Essex, Hearing Research Laboratory (1997).

[19] P. X. Joris, L. H. Carney, P. H. Smith, and T. C. T. Yin, "Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. responses to tones at the characteristic frequency," J. Neurophysiol. **71**, 1022–1036 (1994).

[20] H. Cai, L. H. Carey, and S. H. Colburn, "A model for binaural response properties of inferior colliculus neurons. I. A model with interaural time difference-sensitive excitatory and inhibitory inputs," J. Acoust. Soc. Am. **103**, 475–493 (1998).

[21] T. P. Zahn, "Neural architecture for echo suppression during sound source localization based on spiking neural cell models," Ph.D. thesis, Technische Universität Ilmenau, 2003.

[22] B. Grothe, "Interaction of excitation and inhibition in processing of pure tone and amplitude-modulated stimuli in the medial superior olive of the mustached bat," J. Neurophysiol. **71**, 706–721 (1994).

[23] R. Shannon, F. G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid, "Speech recognition with primarily temporal cues," Science **270**, 303–304 (1995).

[24] D. Bendor and X. Wang, "The neuronal representation of pitch in primate auditory cortex," Nature (London) **436**, 1161–1165 (2005).

[25] P. Friedel, M. Bürck, and J. L. van Hemmen, "Neuronal identification of acoustic signal periodicity," Biol. Cybern. **97**(3), 247–260 (2007).

[26] A. de Cheveigné, "Cancellation model of pitch perception," J. Acoust. Soc. Am. **103**, 1261–1271 (1998).

[27] S. E. Shore, "Recovery of forward-masked responses in ventral cochlear nucleus neurons," Hear. Res. **82**, 31–43 (1995).

[28] P. C. Nelson and L. H. Carney, "A phenomenological model of peripheral and central neural responses to amplitude-modulated tones," J. Acoust. Soc. Am. **116**, 2173–2186 (2004).

[29] M. Dickreiter, *Handbuch der Tonstudiotechnik* [A Handbook of Sound Studio Technology] (Saur, Munich, 1997).

[30] M. J. Hewitt and R. Meddis, "An evaluation of eight computer models of mammalian inner hair-cell function," J. Acoust. Soc. Am. **90**, 904–907 (1991).

Temporal integration and compression near absolute threshold in normal and impaired ears

Christopher J. Plack[a)]

Department of Psychology, Lancaster University, Lancaster, LA1 4YF, United Kingdom

Vicki Skeels

The Hearing Care Centre, 3b Headgate, Colchester, CO1 1NU, United Kingdom

The decrease in absolute threshold with increasing stimulus duration (often referred to as "temporal integration") is greater for listeners with normal hearing than for listeners with sensorineural hearing loss. It has been suggested that the difference is related to reduced basilar-membrane (BM) compression in the impaired group. The present experiment tested this hypothesis by comparing temporal integration and BM compression in normal and impaired ears at low levels. Absolute thresholds were measured for 4, 24, and 44 ms pure-tone signals, with frequencies (f_s) of 2 and 4 kHz. The difference between the absolute thresholds for the 4 and 24 ms signals was used as a measure of temporal integration. Compression near threshold was estimated by measuring the level of a 100 ms off-frequency ($0.45f_s$) pure-tone forward masker required to mask a 44 ms pure-tone signal presented at sensation levels of 5 and 10 dB. There was a significant negative correlation between amount of temporal integration and absolute threshold. However, there was no correlation between absolute threshold and compression at low levels; both normal and impaired ears showed a nearly linear response. The results suggest that the differences in integration between normal and impaired ears cannot be explained by differences in BM compression.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2769829]

I. INTRODUCTION

The lowest detectable level of a sound (absolute threshold) decreases as sound duration increases, for durations up to several hundred milliseconds (see Florentine *et al.*, 1988; Gerken *et al.*, 1990 for a review). This phenomenon is often called "temporal integration" because it is assumed that the ear combines stimulus energy using a long time window in order to improve performance, leading to a lower threshold. However, the effect of duration is also consistent with a model based on the combination of the information in multiple successive short samples or "looks" (Viemeister and Wakefield, 1991), and with models based on the increased probability of a quantum neural spike with increasing duration (Heil and Neubauer, 2003; Krishna, 2002; Meddis, 2006).

The rate of decrease in absolute threshold with increasing duration is generally greater for normal-hearing listeners than for listeners with sensorineural hearing loss (Carlyon *et al.*, 1990; Carlyon and Sloan, 1987; Florentine *et al.*, 1988; Hall and Fernandes, 1983). For normal-hearing listeners the decrease is about 9 dB per decade (tenfold) increase in duration for durations up to about 300 ms, whereas for hearing-impaired listeners the decrease per decade is about 4 dB, although there is considerable variability between listeners, and studies, for both groups (Florentine *et al.*, 1988; Gerken *et al.*, 1990). This variability may be related in part to the

threshold microstructure, the listener-dependent fluctuations in absolute threshold with small frequency changes (Elliott, 1958). It has been shown that the effects of duration for normal-hearing listeners are greater for signals presented in the (low-threshold) valleys of the microstructure than for signals presented at the peaks (Cohen, 1982). Since researchers investigating integration have not generally matched the signal frequency to a particular location in the microstructure for each individual listener, a consequence might be apparent variability between listeners.

The difference between normal-hearing and hearing-impaired listeners in integration over short signal durations is not so clear cut. Oxenham *et al.* (1997) measured masked thresholds for a sinusoidal signal in the presence of a wide-band noise and found no evidence for reduced temporal integration in impaired ears over signal durations from 2 to 10 ms, even for signal levels close to absolute threshold, although there was an impairment-related reduction in integration for longer-duration signals (20–200 ms). If the dependence of absolute threshold on duration is related primarily to the level of the signal, one might expect low sensation-level masked thresholds to produce similar results to absolute thresholds.

It has been suggested that differences in the duration effect may be the result of differences in peripheral compression between normal and impaired ears (Carlyon *et al.*, 1990; Moore, 1991, 1995; Moore and Oxenham, 1998), and that this may be attributable to differences in the response of the basilar membrane (BM). Sensorineural hearing loss is partly a consequence of dysfunction of the outer hair cells (OHCs)

[a)]Author to whom correspondence should be addressed. Electronic mail: c.plack@lancaster.ac.uk

in the organ of Corti. The OHCs are involved in an "active" mechanism that effectively applies gain to stimulation at frequencies close to the characteristic frequency (CF) of each place on the BM (see Yates, 1995). The gain is greatest at low levels and diminishes as level is increased, resulting in a compressive response function (Murugasu and Russell, 1995; Robles et al., 1986; Ruggero et al., 1997; Yates et al., 1990). OHC dysfunction leads to an increase in absolute threshold and a more linear (less compressive) response function (Ruggero and Rich, 1991; Ruggero et al., 1997). Linearization has been observed in hearing-impaired human listeners using behavioral measures (Moore et al., 1999; Nelson et al., 2001; Oxenham and Plack, 1997; Plack et al., 2004).

To enable signal detection, a decrease in duration by some given factor may require an increase in the BM vibration by a given factor. Hence, a given decrease in duration may result in a constant increase in the level of BM vibration [measured, for example, in units of $20\log_{10}(\text{velocity})$] required for signal detection. However, because the BM is compressive in normal ears, a larger change in physical signal level is required to produce this change in the level of BM vibration. The result is a large increase in threshold with decreasing duration. For impaired ears, however, the compression may be reduced or absent. Hence, a smaller change in physical signal level is required, and the effect of duration is reduced. To account for the difference in the duration effect between normal and impaired listeners, a reduction in the compression ratio by a factor of about 2 is required, corresponding to a factor of 2 reduction in the duration effect.

As noted by Oxenham et al. (1997), a potential problem with the BM compression hypothesis is that, at low signal levels, the gain of the active mechanism does not vary greatly with level, and hence the BM response function may be close to linear even in normal ears (Murugasu and Russell, 1995; Ruggero et al., 1997). If a dysfunction of the OHCs does not produce a significant change in compression at low levels, then it is not clear how the hypothesis can account for the factor of 2 difference in the effect of duration on absolute threshold. However, absolute threshold for short signals may be sufficiently high in some cases to be within the more compressive region of the BM response function (although the results of Oxenham et al., 1997 suggest that the impairment-related difference in integration may be reduced for short signals, exactly the opposite prediction of the BM hypothesis). Furthermore, there have been very few estimates of the human BM response close to absolute threshold. These measures have used behavioral techniques interpreted to reflect aspects of the BM response. There is some inconsistency in the findings for levels in the region of 25–35 dB sound pressure level (SPL), with some studies suggesting a nearly linear response (Nelson et al., 2001; Oxenham and Plack, 1997; Yasin and Plack, 2003) and some suggesting residual compression (Lopez-Poveda et al., 2003; Nelson and Schroder, 2004).

The aim of the present experiment was to estimate temporal integration and BM compression over the *same range of levels* in each listener, for both normal-hearing and hearing-impaired listeners, in order to test the BM compression hypothesis. If the duration effect is dependent on BM compression, then the two measures should be highly correlated.

II. METHOD

A. Rationale

Over recent years, several behavioral techniques have been developed to estimate the response of the human BM (Hicks and Bacon, 1999; Nelson et al., 2001; Oxenham and Plack, 1997; Plack and Drga, 2003; Plack and O'Hanlon, 2003; Plack et al., 2006). Most of these techniques have involved comparisons of the effects of forward maskers at and below the signal frequency. Forward masking, rather than simultaneous masking, is used to avoid simultaneous interactions on the BM (e.g., suppression) that complicate the interpretation of the results (Oxenham and Plack, 1997, 1998). Since the BM response at a given site to a masker with a frequency well below the CF of this site is linear (Ruggero et al., 1997), the off-frequency masking function can be used as a linear reference to derive the BM response to a tone at CF. The technique used here is based on that of Oxenham and Plack (1997). The 44 ms pure-tone signal was presented at either 5 or 10 dB sensation level. These levels were chosen so that, for a given listener, the compression estimate was made over approximately the same range of sound pressure levels as those corresponding to absolute threshold for the 4 and 24 ms signals used to estimate temporal integration. For each signal level, the level of a lower-frequency forward masker was found that was sufficient to mask the signal. Because the response to the masker is linear at the signal place, the difference in the masker levels is an estimate of the growth in BM excitation corresponding to the 5 dB growth in signal level (although a correction has to be added for the internal noise floor, see Sec. III B).

B. Stimuli and equipment

The pure-tone signal had a frequency (f_s) of either 2 or 4 kHz, and a total duration of either 4, 24, or 44 ms (including 2 ms raised-cosine onset and offset ramps). The pure-tone masker had a frequency of $0.45f_s$ (i.e, either 900 or 1800 Hz) and a total duration of 200 ms (including 5 ms raised-cosine ramps). The onset ramp of the signal commenced immediately after the end of the offset ramp of the masker (0 ms masker-signal gap).

The experiment was run using custom-made software on a PC workstation located outside a double-walled sound-attenuating booth. Stimuli were digitally generated and were produced using a RME Digi96/8 PAD 24 bit soundcard set at a clocking rate of 48 kHz. The soundcard included an anti-aliasing filter. The headphone output of the soundcard was fed via a patch panel in the sound booth wall, without filtering or amplification, to Sennheiser HD 580 circumaural headphones. Each listener sat in the booth and decisions were recorded via a computer keyboard. Listeners viewed a computer monitor through a window in the sound booth.

TABLE I. Audiometric thresholds (in dB hearing level) for the tested ears of six of the seven hearing-impaired listeners.

Listener	Age	Ear	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
I2	60	R	40	40	40	35	70	75
I3	60	R	55	55	55	50	45	55
I4	63	R	30	20	25	45	65	80
I5	74	L	25	25	20	45	55	60
I6	74	R	30	30	35	45	45	65
I7	79	L	55	70	70	60	60	60

Lights on the monitor display flashed on and off concurrently with each stimulus presentation and provided feedback at the end of each trial.

C. Procedure

All stimuli were presented monaurally. Normal-hearing listeners were tested in the right ear. Hearing-impaired listeners were tested in either the left or right ear (see Sec. II D). The experiment used a three-interval, three-alternative, forced-choice adaptive tracking procedure with a 300 ms inter-stimulus interval. In the first phase of the experiment (absolute thresholds), one of the intervals (chosen at random) contained the signal and the other two intervals were silent. The signal level was varied using a two-down one-up rule to obtain the level needed to achieve 70.7% correct (Levitt, 1971). The step size of the adaptive track was 4 dB for the first four turnpoints and 2 dB for 12 subsequent turnpoints. The threshold estimate for each block was taken as the mean signal level over the last 12 turnpoints. Signal durations of 4, 24, and 44 ms were tested in separate blocks, and the order of the conditions was randomized. After an initial training procedure over one or two blocks, threshold estimates were made over four blocks for each condition. To reduce the effects of outliers and to improve reliability, the most deviant threshold (that furthest from the mean) was discarded, and the final estimate of threshold was taken as the mean of the remaining three.

In the second phase of the experiment (off-frequency masker thresholds), all three intervals contained the lower-frequency forward masker and one of the intervals (chosen at random) contained the 44 ms signal at either 5 or 10 dB sensation level. The sound pressure levels were determined separately for each listener from the thresholds in the first phase. The masker level was varied adaptively using a two-up one-down rule. The step sizes, number of turnpoints, and method of estimating the final threshold masker level, were the same as those for the signal thresholds in the first phase.

Signal and masker thresholds at a given frequency were measured in the same 2 h session. The 2 and 4 kHz conditions were tested in separate sessions.

D. Listeners

Six normal-hearing (ages 21–39) and seven hearing-impaired listeners (ages 42–79) were tested. The criterion for hearing impairment was an absolute threshold greater than 20 dB SPL for the 44 ms signal at the test frequency. The ages and audiometric thresholds of six of the seven hearing-impaired listeners are given in Table I. I1 (age 42) was only mildly impaired in the tested (left) ear, and audiometric data are not available for this listener. The degree of impairment varied from mild to moderate at the test frequencies. With the exception of I1, the hearing-impaired listeners were recruited from the Colchester Hearing Care Centre, and had been diagnosed previously with sensorineural hearing loss on the basis of elevated audiometric thresholds, the absence of an air-bone gap, and evidence of loudness recruitment. The ear and test frequencies for each impaired listener were chosen with the criterion that audiometric thresholds at frequencies an octave above and an octave below the test frequency were no more than 20 dB below that at the test frequency, hence minimizing the potential effects of spectral splatter (Hall and Fernandes, 1983). Four of the normal-hearing listeners were tested at both frequencies, but due to time constraints with the hearing-impaired listeners, only one of these listeners was tested at both frequencies. For each frequency, therefore, measurements were made for five normal-hearing listeners and for four hearing-impaired listeners (as indicated in Table II). One of the normal-hearing listeners was author C.P. The other listeners were paid an hourly wage for their participation.

III. RESULTS AND ANALYSES

A. Temporal integration

Absolute thresholds for the three signal durations at each frequency are presented in Table II. Fig. 1(a) shows the difference between the absolute thresholds for the 4 and 24 ms signals (a measure of temporal integration) plotted against the absolute thresholds for the 44 ms signal. The straight line shows a linear regression for the combined data (it was deemed reasonable to combine the data in this way, since the mean absolute thresholds for the normal-hearing listeners at the two frequencies differed by only 1.3 dB). Although there is considerable variability, consistent with previous studies (Florentine *et al.*, 1988; Gerken *et al.*, 1990), the data show the expected decrease in temporal integration with increasing absolute threshold. Correlation tests (two-tailed) using Pearson's r revealed significant negative correlations between threshold difference and absolute threshold at 4 kHz $[r(7)=-0.71, p=0.032]$, and for the combined data $[r(16)=-0.51, p=0.032]$, but not at 2 kHz $[r(7)=-0.38, p=0.32]$. The mean threshold difference is 8.8 dB for the normal-hearing group and 5.9 dB for the hearing-impaired group, and the difference between these means is significant two-tailed $[t(16)=2.20, p=0.043]$. The increase in energy of the signal as duration is increased from 4 to 24 ms (taking into

TABLE II. Individual absolute thresholds for the three signal durations, and masker levels at threshold for the two signal levels, at each frequency. Standard errors are given in parentheses.

	2 kHz					4 kHz				
	Absolute Threshold (dB SPL)			Masker Level (dB SPL)		Absolute Threshold (dB SPL)			Masker Level (dB SPL)	
Listener (Ear)	4 ms	24 ms	44 ms	5 dB SL	10 dB SL	4 ms	24 ms	44 ms	5 dB SL	10 dB SL
N1 (R)	17.2 (0.4)	5.9 (0.1)	3.3 (0.2)	72.2 (0.5)	80.5 (1.0)	11.8 (0.3)	3.3 (0.8)	2.4 (0.4)	66.3 (0.7)	72.5 (0.2)
N2 (R)	22.2 (1.0)	11.8 (0.9)	7.1 (0.2)	71.9 (0.3)	78.1 (0.1)	20.6 (0.3)	13.9 (0.4)	12.9 (0.2)	68.0 (0.1)	75.5 (0.2)
N3 (R)	19.5 (1.3)	9.1 (0.5)	7.4 (0.3)	70.9 (0.5)	77.0 (0.0)	16.5 (0.3)	7.5 (0.2)	6.5 (0.3)	76.7 (0.1)	82.4 (0.3)
N4 (R)	25.9 (1.0)	12.0 (0.5)	11.8 (0.5)	66.3 (0.8)	71.7 (0.2)	18.8 (0.8)	9.4 (0.4)	9.7 (1.6)	72.5 (1.6)	83.7 (0.3)
N5 (R)	21.0 (0.8)	18.0 (0.6)	13.7 (0.4)	68.2 (0.4)	74.7 (0.4)					
N6 (R)						24.1 (1.2)	18.9 (0.5)	18.3 (0.5)	74.4 (0.5)	80.5 (0.3)
I1 (L)	32.4 (0.8)	25.8 (1.2)	24.8 (0.8)	68.4 (0.2)	73.4 (0.3)					
I2 (R)	48.0 (0.4)	40.5 (0.2)	39.8 (0.5)	72.5 (0.3)	77.1 (0.2)					
I3 (R)	57.3 (0.4)	52.6 (0.5)	51.1 (0.2)	61.8 (0.2)	71.2 (1.5)	54.8 (0.3)	50.3 (0.8)	48.1 (0.1)	79.6 (0.6)	83.9 (0.1)
I4 (R)	66.8 (1.0)	57.1 (0.3)	55.5 (0.2)	69.5 (0.2)	75.8 (0.2)					
I5 (L)						48.7 (0.0)	42.7 (0.5)	42.1 (0.7)	68.6 (1.2)	79.7 (1.2)
I6 (R)						54.9 (0.5)	53.1 (1.3)	45.3 (3.3)	92.4 (0.7)	96.3 (0.40)
I7 (L)						69.3 (0.1)	63.5 (0.2)	62.0 (0.4)	73.7 (0.3)	82.8 (0.1)

account the ramps) is 11.6 dB. Assuming that threshold in dB is linearly related to the logarithm of signal energy (and hence to the logarithm of the effective duration), these values need to be multiplied by 10/11.6 to obtain the per-decade equivalent. Hence, the amount of integration is equivalent to about 7.6 dB per decade for the normal-hearing group and about 5.1 dB per decade for the hearing-impaired group.

Between 24 and 44 ms, the mean decrease in threshold is similar for the two groups: 1.7 dB for the normal-hearing group and 2.1 dB for the hearing-impaired group. As can be seen in Table II, there is considerable variability between listeners, and because the energy increase is not large (2.9 dB), a comparison of thresholds over these longer durations is not a sensitive measure of temporal integration.

B. Compression

Masker levels at threshold for the two signal levels at each frequency are presented in Table II. Figure 1(b) shows the difference between the masker thresholds for the 5 and 10 dB sensation level signals (a measure of compression) plotted against the absolute thresholds for the 44 ms signal. There is little variation in the masker-level difference with absolute threshold. Correlation tests using Pearson's r revealed no significant correlations between masker-level difference and absolute threshold at either 2 kHz [$r(7)=0.13$, $p=0.75$], 4 kHz [$r(7)=-0.012$, $p=0.98$], or for the combined

data [$r(16)=0.052$, $p=0.84$]. The mean masker-level difference is 6.9 dB for the normal-hearing group and 6.7 dB for the hearing-impaired group.

It is possible to convert these values into an estimate of the compression exponent, c, by simply dividing the masker-level difference by the difference in signal levels. This gives exponent estimates of 1.4 for the normal-hearing group and 1.3 for the hearing-impaired group, suggestive of an expansive nonlinearity. However, this does not take into account the possibility that the detectability of the signal was influenced by the proximity of absolute threshold. In the past this has been modeled using an "internal noise floor" that is assumed to determine absolute threshold (Humes and Jesteadt, 1989; Oxenham and Moore, 1995; Plack and O'Hanlon, 2003; Plack and Oxenham, 1998). The internal noise acts as another masker, and reduces the forward masker level required to mask the signal. The effect of internal noise will be greater for the lower-level (5 dB sensation level) signal, and hence the masker level at threshold will be reduced compared to that for the 10 dB sensation-level signal.

To take account of the proximity of absolute threshold in the calculation of c, the following derivation was used, based on that employed in previous studies (Plack and O'Hanlon, 2003; Plack and Oxenham, 1998; Plack et al., 2006). First, it was assumed that the ratio of internal signal intensity to the combined (off-frequency masker plus noise floor) internal masker intensity is a constant at threshold. This assumption
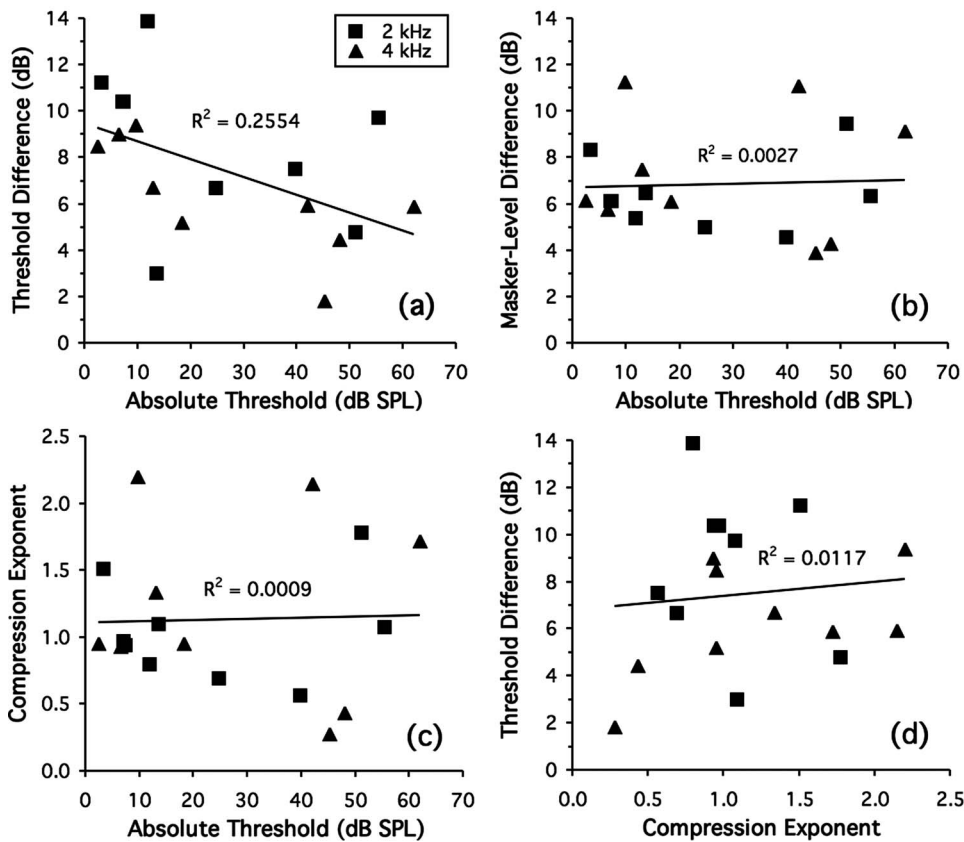
FIG. 1. Scatterplots illustrating the relations between the variables measured in the experiment. Each point represents the response of a single listener/frequency combination. Panel (a): the difference in absolute threshold between the 4 and 24 ms signals plotted against absolute threshold for the 44 ms signal. Panel (b): the difference between the levels of an off-frequency masker required to mask a 5 or 10 dB sensation level 44-ms signal plotted against absolute threshold for the 44 ms signal. Panel (c): the derived compression exponents plotted against absolute threshold for the 44 ms signal. Panel (d): the difference in absolute threshold between the 4 and 24 ms signals plotted against the derived compression exponents. The straight lines on the plots show linear regressions, and the R^2 values are provided on the plots.

is consistent with results showing a linear growth of forward masking when both masker and signal are subject to the same BM compression (Oxenham and Plack, 1997; Plack and Oxenham, 1998), and a linear growth of forward masking for hearing-impaired listeners (Oxenham and Moore, 1995; Oxenham and Plack, 1997). It was also assumed that the internal signal intensity is a power-law transformation of physical signal intensity. Hence

$$I = k_1 S^c, \tag{1}$$

where I is the internal masker intensity, S is the physical signal intensity (average intensity during steady-state portion) at masked threshold, c is the compression exponent, and k_1 is a constant with respect to masker and signal intensity, but dependent on other characteristics of the signal such as signal duration (if the signal duration is increased, k_1 will increase to account for the increase in masker level at threshold). I is comprised of two components. First, the off-frequency masker intensity, M, assumed to be linearly related to its effective internal masker intensity by the constant k_2; k_2 depends on the frequency and temporal separation of masker and signal, and on the temporal characteristics of the masker. The assumption is consistent with the linear growth of off-frequency forward masking observed for hearing-impaired listeners (Moore et al., 1999; Oxenham and Plack, 1997). k_1 and k_2 accommodate linear processes often modeled using a temporal weighting function such as the temporal window (e.g. Plack and Oxenham, 1998). Since the temporal and spectral characteristics of the masker and signal are fixed, we do not need to make such processes explicit in the present analysis. The second component of I is the internal

intensity-like quantity associated with the noise floor (N). Hence

$$I = k_2 M + N. \tag{2}$$

N can be expressed in terms of the compressed signal intensity using Eq. (1)

$$N = k_1 S_0^c, \tag{3}$$

where S_0 is the signal intensity at absolute threshold. Substituting into Eqs. (1) and (2) and solving for k_1 gives

$$k_1 = \frac{k_2 M}{S^c - S_0^c}. \tag{4}$$

Since k_1 and k_2 are constant with respect to level

$$\frac{M_5}{S_5^c - S_0^c} = \frac{M_{10}}{S_{10}^c - S_0^c}, \tag{5}$$

where S_5 and S_{10} are the signal intensities corresponding to 5 and 10 dB sensation levels respectively, and M_5 and M_{10} are the off-frequency masker intensities required to mask these signals. Since all the values except c are known, it is possible to compute the value of c. This was achieved using the "solver" routine in Microsoft Excel.

Figure 1(c) shows the compression exponents plotted against the absolute thresholds for the 44 ms signal. There is little variation in compression with absolute threshold, and none of the correlations are significant. The compression exponents derived from the mean masker levels are 1.2 for the normal-hearing group and 1.1 for the hearing-impaired group, showing nearly linear response growth in both cases.

C. J. Plack and V. Skeels: Temporal integration and compression

Figure 1(d) shows the threshold differences between the 4 and 24 ms tones plotted against the compression exponents. There are no significant correlations between the compression exponents and the amount of temporal integration, at either 2 kHz [$r(7)=-0.21$, $p=0.60$], 4 kHz [$r(7)=0.53$, $p=0.14$], or for the combined data [$r(16)=0.11$, $p=0.67$]. The strongest correlation, at 4 kHz, is actually in the wrong direction to account for the variation in temporal integration in terms of the BM compression hypothesis, showing an increase in temporal integration with decreasing compression. Crucially, the correlation between temporal integration and absolute threshold is significant at 4 kHz [$r(6)=-0.81$, $p=0.015$], and for the combined data [$r(15)=-0.51$, $p=0.036$], even when the effects of the compression exponent are removed using partial correlation. This demonstrates that at least part of the variation in temporal integration with absolute threshold cannot be explained by low-level compression.

IV. DISCUSSION

A. Comparison with previous studies

The mean difference in absolute threshold between the 4 and 24 ms signals is 8.8 dB for the normal-hearing group and 5.9 dB for the hearing-impaired group, equivalent to 7.6 and 5.1 dB per decade, respectively. Hence, the difference for the normal-hearing listeners is slightly less than would be predicted by a perfect energy integrator, but within the range of previous data (Florentine *et al.*, 1988; Gerken *et al.*, 1990). The amount of temporal integration for the hearing-impaired listeners is greater than has been reported in some studies (Carlyon *et al.*, 1990; Florentine *et al.*, 1988), possibly because some of the listeners categorized as impaired in the present study had only a mild loss at the frequency tested. Furthermore, in the present study integration was measured over relatively short durations, and there is some evidence that the difference between normal and impaired ears is much reduced at short durations (Oxenham *et al.*, 1997).

It should be noted that there is a potential confound in most studies of temporal integration and hearing loss, including the present study, in that the hearing-impaired listeners have usually been older than the normal-hearing listeners used for comparison. An age-related impairment has been reported for some temporal tasks, including duration discrimination (Fitzgibbons and Gordon-Salant, 1994) and gap detection (Pichora-Fuller *et al.*, 2006). However, Carlyon and Sloan (1987) report large impairment-related differences in temporal integration between the ears of listeners with unilateral or asymmetric loss. Analysis of the data from Table I of their study reveals a strong negative correlation between integration and absolute threshold at the test frequency [$r(12)=-0.91$, $p<0.0005$], but a weak positive (i.e., in the wrong direction) correlation between integration and age [$r(12)=0.02$, $p=0.94$]. These results suggest that hearing loss, rather than age, is the main factor influencing the duration effect.

There have been very few estimates of BM compression in humans at levels this close to absolute threshold. Using the same measure of growth of off-frequency masking, Oxenham and Plack (1997) estimated BM response growth down to 30 dB SPL. The slope of the mean response function between 30 and 35 dB SPL was 0.95 dB/dB. Studies using the temporal masking curve (TMC) technique, in which the threshold levels of on- and off-frequency maskers are measured as a function of masker-signal interval, have reported compression exponents down to levels as low as 25 dB SPL. Values range from about 0.5–0.6 (Lopez-Poveda *et al.*, 2003; Nelson and Schroder, 2004) to about 0.8–1.0 (Nelson *et al.*, 2001; Yasin and Plack, 2003). Finally, using a technique based on additivity of masking, Plack and O'Hanlon (2003) estimated a mean compression exponent of 0.96 for a 10 ms signal at a sensation level of 10 dB. Overall, the bulk of the data seem to confirm that the normal BM response function is roughly linear at low levels, although there are some exceptions in the TMC literature.

Despite the present findings, it seems likely that BM compression has a major influence on temporal integration for signal levels well above absolute threshold. Oxenham *et al.* (1997) reported a large increase in the duration effect for sinusoids in noise maskers at medium levels compared to low and high levels, for normal-hearing listeners. This is consistent with the peak in BM compression at mid levels observed in some behavioral experiments (Lopez-Poveda *et al.*, 2003; Nelson *et al.*, 2001; Oxenham and Plack, 1997). In contrast, the hearing-impaired listeners in their study did not show any variation in integration with level, consistent with the expected loss of BM compression.

B. Why is there a difference in temporal integration between normal and impaired ears?

1. Compression in the inner hair cell

The present data provide evidence against the hypothesis that the difference in temporal integration between normal and impaired ears is a consequence of differences in compression at the level of the BM. However, it is possible that there is a difference in compression between normal and impaired ears at a site central to the BM, possibly at the level of the inner hair cell (IHC). Compression occurring after BM filtering will likely be independent of the frequency of stimulation for a given frequency channel. Hence, such compression will not be observable using the off-frequency masking technique described here, although it may still be revealed by the additivity technique (Plack and O'Hanlon, 2003; Plack *et al.*, 2006). Consistent with this hypothesis is the demonstration by Carlyon *et al.* (1990) that the psychometric function of detectability against signal level near absolute threshold for brief signals is steeper for hearing-impaired listeners than for normal-hearing listeners. This could be due to differences in compression, although, as Carlyon *et al.* point out, differences in the growth of standard deviation with level will also create differences between the psychometric functions. A difference in post-BM compression might explain the puzzling finding that the slopes of TMCs (plots of masker threshold level against masker-signal interval) are steeper for normal than for impaired ears at high signal frequencies, even when the masker is much lower in frequency than the signal

(Lopez-Poveda *et al.*, 2005; Plack *et al.*, 2004; Rosengard *et al.*, 2005). A masker much lower in frequency than the signal is usually assumed to be processed linearly by the BM at the signal place. The difference in slopes could be explained by greater compression in the normal ears at a site central to the BM, possibly at the level of the IHC (Lopez-Poveda *et al.*, 2005). As noted by Lopez-Poveda *et al.* (2005), some reports suggest that the growth in IHC receptor potential with level is compressive even for tones well below CF (Cheatham and Dallos, 2001; Patuzzi and Sellick, 1983). The IHC compression ratio of approximately 2:1 matches the ratio of off-frequency TMC slopes between normal and impaired ears. If dysfunction of the IHCs due to hearing loss leads to a complete loss of this compression, then the amount of temporal integration (in dB per decade) might also be expected to be reduced by a factor of 2, which is at least broadly consistent with the data.

2. Pressure integration in the inner hair cell

Another possible explanation for the effects of impairment on temporal integration, also involving mechanisms at the IHC level, was proposed recently by Neubauer and Heil (2004). In an earlier paper (Heil and Neubauer, 2003) the authors argued that absolute threshold depends on the integration of the stimulus pressure envelope at the IHC synapse. To account for the effects of sensorineural hearing loss they suggest that IHC dysfunction raises the base line above which sound pressure is effective in driving the system. Hence, a given change in stimulus level will produce a larger proportional change in effective pressure for the impaired IHC. For the same integration time constant, this results in a more gradual decline in threshold level (in dB) over time (Neubauer and Heil, 2004). The hypothesis does not depend on a reduction in IHC compression, but instead implies an impairment-related nonlinearity in the IHC equivalent to expansion. The model of Neubauer and Heil predicts the steeper psychometric functions for hearing-impaired listeners observed by Carlyon *et al.* (1990). The model can also account for the smaller difference in integration between normal and impaired ears at short durations (Oxenham *et al.*, 1997), since the effect of the elevated base line decreases as amplitude increases. Although simulations using the model have not been attempted here, the present data are consistent with this explanation, and support the contention of Neubauer and Heil that the effects of impairment on integration cannot be explained in terms of differences in BM compression.

3. Threshold microstructure

Finally, there may be a role of threshold microstructure. As described in the Introduction, the effects of duration for normal-hearing listeners are greater for signals presented in the valleys of the microstructure than for signals presented at the peaks (Cohen, 1982). For example, for a duration increase of 8–32 ms (durations included 4 ms linear ramps) Cohen reports a threshold decrease of about 8 dB for a valley location, but only about 4 dB for a peak location. Cohen explains these data in terms of the spectral spread of the signal at short durations. For a peak location, the signal energy will spread to the neighboring (high-sensitivity) valleys, reducing threshold. For a valley location, the signal energy will spread to neighboring (low-sensitivity) peaks, increasing threshold. These findings are relevant to the present study since the fine structure variations in threshold with frequency, and the associated otoacoustic emission fine structure, may be reduced in cases of cochlear dysfunction (for a review, see Mauermann *et al.*, 2004). Assuming that the frequencies chosen for the present study were randomly located with respect to features in the fine structure, one might therefore expect a greater variation in the 4–24 ms threshold difference for the normal-hearing listeners than for the hearing-impaired listeners. This was observed, although the difference in variance was not significant [$F(9,7)=1.82, p > 0.05$].

The t test described in Sec. III A seems to confirm that the correlation with absolute threshold was not driven entirely by differences in the normal-hearing group; there was a significant difference in the duration effect between the two groups. Hence, for differences in threshold microstructure to account for the present results, the effects on temporal integration of randomly locating the signal frequency relative to the microstructure should be greater when the microstructure is more pronounced. However, on the basis that integration is dependent on the interaction of microstructure with spectral splatter, it seems probable that the reverse will be the case. The effects of splatter on reducing threshold for a peak location are likely to be greater than the effects on increasing threshold for a valley location. This was confirmed with a simulation conducted assuming a sinusoidal variation in microstructure threshold pressure with log CF, with a period of 0.05 CF and a peak-to-valley depth of 10 dB. The simulated microstructure (expressed as equivalent attenuation in units of power) was multiplied by the spectra of the 4 and 24 ms 2 kHz signals used in the present study, and the resulting power spectrum integrated across frequency. In comparison with a flat microstructure, a peak location decreased the energy difference between the two signals by 3.0 dB, whereas a valley location increased the energy difference by just 1.8 dB. Hence, the average threshold difference may be *reduced* when the microstructure is pronounced (although this is a very tentative analysis that depends on the form of the microstructure variations). Furthermore, the results of Carlyon *et al.* (1990) suggest that the interaction between spectral splatter and the microstructure cannot be a complete explanation for the difference between normal and impaired ears. Instead of continuous tones, Carlyon *et al.* used trains of tone pulses designed to minimize the spectral differences between long and short durations. They found an effect of hearing loss on temporal integration similar to that of previous studies.

Another contributory effect of threshold microstructure on temporal integration was suggested by Mauermann *et al.* (2004). It is thought that the microstructure may result from the constructive or destructive interference of the original traveling wave with reflections between the stimulus place and the oval window (Talmadge *et al.*, 1998). Mauermann *et al.* suggest that the microstructure variations in threshold are

C. J. Plack and V. Skeels: Temporal integration and compression

reduced at short durations because there is not enough time for the interference pattern to build up. This will tend to reduce the effects of duration for peak placements relative to valley placements. For this process to account for the difference between the normal and impaired groups observed in the present study, we would have to assume that the overall net effect (averaging valley and peak placements) of the buildup of microstructure is to reduce threshold. It is not clear that this would be the case.

Furthermore, neither of the explanations based on microstructure can account for the fact that a difference between normal and impaired ears has been observed for duration changes spanning much longer durations (>20 ms, see, for example, Florentine *et al.*, 1988), long enough for splatter effects to be negligible, and long enough for the interference pattern to build up. Indeed, as stated earlier, the difference between the two groups may be reduced at short durations (Oxenham *et al.*, 1997), a finding in opposition to the prediction of these hypotheses.

V. CONCLUSIONS

(i)   Temporal integration, as estimated by the difference in absolute threshold between 4 and 24 ms pure tones at 2 and 4 kHz, decreases with increasing hearing loss.

(ii)   There is little effect of hearing loss on estimates of BM compression made over the same range of levels. For both normal and impaired ears the BM response is close to linear at low levels.

(iii)   The difference in temporal integration at low levels between normal-hearing and hearing-impaired listeners cannot be explained by differences in BM compression.

Carlyon, R. P., Buus, S., and Florentine, M. (**1990**). "Temporal integration of trains of tone pulses by normal and by cochlearly impaired listeners," J. Acoust. Soc. Am. **87**, 260–268.

Carlyon, R. P., and Sloan, E. P. (**1987**). "The overshoot effect and sensory hearing impairment," J. Acoust. Soc. Am. **82**, 1078–1081.

Cheatham, M. A., and Dallos, P. (**2001**). "Inner hair cell response patterns: Implications for low-frequency hearing," J. Acoust. Soc. Am. **110**, 2034–2044.

Cohen, M. F. (**1982**). "Detection threshold microstructure and its effect on temporal integration data," J. Acoust. Soc. Am. **71**, 405–409.

Elliott, E. (**1958**). "A ripple effect in the audiogram," Nature (London) **181**, 1076.

Fitzgibbons, P. J., and Gordon-Salant, S. (**1994**). "Age effects on measures of auditory duration discrimination," J. Speech Hear. Res. **37**, 662–670.

Florentine, M., Fastl, H., and Buus, S. (**1988**). "Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking," J. Acoust. Soc. Am. **84**, 195–203.

Gerken, G. M., Bhat, V. K. H., and Hutchinson-Clutter, M. (**1990**). "Auditory temporal integration and the power function model," J. Acoust. Soc. Am. **88**, 767–778.

Hall, J. W., and Fernandes, M. A. (**1983**). "Temporal integration, frequency resolution, and off-frequency listening in normal-hearing and cochlearly impaired listeners," J. Acoust. Soc. Am. **74**, 1172–1177.

Heil, P., and Neubauer, H. (**2003**). "A unifying basis of auditory thresholds based on temporal summation," Proc. Natl. Acad. Sci. U.S.A. **100**, 6151–6156.

Hicks, M. L., and Bacon, S. P. (**1999**). "Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing," J. Acoust. Soc. Am. **105**, 326–338.

Humes, L. E., and Jesteadt, W. (**1989**). "Models of the additivity of masking," J. Acoust. Soc. Am. **85**, 1285—1294.

Krishna, B. S. (**2002**). "A unified mechanism for spontaneous-rate and first-spike timing in the auditory nerve," J. Comput. Neurosci. **13**, 17–91.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Lopez-Poveda, E. A., Plack, C. J., and Meddis, R. (**2003**). "Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing," J. Acoust. Soc. Am. **113**, 951–960.

Lopez-Poveda, E. A., Plack, C. J., Meddis, R., and Blanco, J. L. (**2005**). "Cochlear compression between 500 and 8000 Hz in listeners with moderate sensorineural hearing loss," Hear. Res. **205**, 172–183.

Mauermann, M., Long, G. R., and Kollmeier, B. (**2004**). "Fine structure of hearing threshold and loudness perception," J. Acoust. Soc. Am. **116**, 1066–1080.

Meddis, R. (**2006**). "Reply to comment on Auditory-nerve first-spike latency and auditory absolute threshold: A computer model," J. Acoust. Soc. Am. **120**, 1192–1193.

Moore, B. C. J. (**1991**). "Characterization and simulation of impaired hearing: Implications for hearing aid design," Ear Hear. **12**, 154S–161S.

Moore, B. C. J. (**1995**). *Perceptual Consequences of Cochlear Damage* (OUP, Oxford).

Moore, B. C. J., and Oxenham, A. J. (**1998**). "Psychoacoustic consequences of compression in the peripheral auditory system," Psychol. Rev. **105**, 108–124.

Moore, B. C. J., Vickers, D. A., Plack, C. J., and Oxenham, A. J. (**1999**). "Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism," J. Acoust. Soc. Am. **106**, 2761–2778.

Murugasu, E., and Russell, I. J. (**1995**). "Salicylate ototoxicity: The effects on basilar membrane displacement, cochlear microphonics, and neural responses in the basal turn of the guinea pig cochlea," Aud. Neurosci. **1**, 139–150.

Nelson, D. A., and Schroder, A. C. (**2004**). "Peripheral compression as a function of stimulus level and frequency region in normal-hearing listeners," J. Acoust. Soc. Am. **115**, 2221–2233.

Nelson, D. A., Schroder, A. C., and Wojtczak M. (**2001**). "A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 2045–2064.

Neubauer, H., and Heil, P. (**2004**). "Towards a unifying basis of auditory thresholds: The effects of hearing loss on temporal integration reconsidered," J. Assoc. Res. Otolaryngol. **5**, 436–458.

Oxenham, A. J., and Moore, B. C. J. (**1995**). "Additivity of masking in normally hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **98**, 1921–1934.

Oxenham, A. J., Moore, B. C. J., and Vickers, D. A. (**1997**). "Short-term temporal integration: Evidence for the influence of peripheral compression," J. Acoust. Soc. Am. **101**, 3676–3687.

Oxenham, A. J., and Plack, C. J. (**1997**). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **101**, 3666–3675.

Oxenham, A. J., and Plack, C. J. (**1998**). "Suppression and the upward spread of masking," J. Acoust. Soc. Am. **104**, 3500–3510.

Patuzzi, R., and Sellick, P. M. (**1983**). "A comparison between basilar membrane and inner hair cell receptor potential input-output functions in the guinea pig cochlea," J. Acoust. Soc. Am. **74**, 1734–1741.

Pichora-Fuller, K., Schneider, B. A., Benson, N. J., Hamstra, S. J., and Storzer, E. (**2006**). "Effect of age on detection of gaps in speech and nonspeech markers varying in duration and spectral symmetry," J. Acoust. Soc. Am. **119**, 1143–1155.

Plack, C. J., and Drga, V. (**2003**). "Psychophysical evidence for auditory compression at low characteristic frequencies," J. Acoust. Soc. Am. **113**, 1574–1586.

Plack, C. J., Drga, V., and Lopez-Poveda, E. A. (**2004**). "Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss," J. Acoust. Soc. Am. **115**, 1684–1695.

Plack, C. J., and O'Hanlon, C. G. (**2003**). "Forward masking additivity and

auditory compression at low and high frequencies," J. Assoc. Res. Oto-laryngol. **4**, 405–415.

Plack, C. J., and Oxenham, A. J. (**1998**). "Basilar-membrane nonlinearity and the growth of forward masking," J. Acoust. Soc. Am. **103**, 1598–1608.

Plack, C. J., Oxenham, A. J., and Drga, V. (**2006**). "Masking by inaudible sounds and the linearity of temporal summation," J. Neurosci. **26**, 8767–8773.

Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). "Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and phase responses," J. Acoust. Soc. Am. **80**, 1364–1374.

Rosengard, P. S., Oxenham, A. J., and Braida, L. D. (**2005**). "Comparing different estimates of cochlear compression in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **117**, 3028–3041.

Ruggero, M. A., and Rich, N. C. (**1991**). "Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane," J. Neurosci. **11**, 1057–1067.

Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **101**, 2151–2163.

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.

Viemeister, N. F., and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865.

Yasin, I., and Plack, C. J. (**2003**). "The effects of a high-frequency suppressor on tuning curves and derived basilar membrane response functions," J. Acoust. Soc. Am. **114**, 322–332.

Yates, G. K. (**1995**). "Cochlear structure and function," in *Hearing*, B. C. J. Moore ed. (Academic, San Diego), pp. 41–73.

Yates, G. K., Winter, I. M., and Robertson, D. (**1990**). "Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range," Hear. Res. **45**, 203–220.

Hearing loss from interrupted, intermittent, and time varying non-Gaussian noise exposure: The applicability of the equal energy hypothesis

Roger P. Hamernik,[a)] Wei Qiu, and Bob Davis
Auditory Research Laboratory, State University of New York at Plattsburgh, 107 Beaumont Hall, Plattsburgh, New York 12901

Sixteen groups of chinchillas ($N=140$) were exposed to various equivalent energy noise paradigms at 100 dB(A) or 103 dB(A) SPL. Eleven groups received an interrupted, intermittent, and time varying (IITV) non-Gaussian exposure quantified by the kurtosis statistic. The IITV exposures, which lasted for 8 h/day, 5 days/week for 3 weeks, were designed to model some of the essential features of an industrial workweek. Five equivalent energy reference groups were exposed to either a Gaussian or non-Gaussian 5 days, 24 h/day continuous noise. Evoked potentials were used to estimate hearing thresholds and surface preparations of the organ of Corti quantified the sensory cell population. For IITV exposures at an equivalent energy and kurtosis, the temporal variations in level did not alter trauma and in some cases the IITV exposures produced results similar to those found for the 5 day continuous exposures. Any increase in kurtosis at a fixed energy was accompanied by an increase in noise-induced trauma. These results suggest that the equal energy hypothesis is an acceptable approach to evaluating noise exposures for hearing conservation purposes provided that the kurtosis of the amplitude distribution is taken into consideration. Temporal variations in noise levels seem to have little effect on trauma. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2775160]

I. INTRODUCTION

Industrial and military noise exposures are most often interrupted and intermittent with sound levels that vary over the course of a work shift (Taylor *et al.*, 1984). The presence of transients, either impacts or noise bursts, further complicates the evaluation of these noise environments for hearing conservation purposes. The hearing loss that results from such exposures usually accumulates over many years or decades of exposure. Much of our understanding of the effects of noise on hearing is derived from acute animal model experiments that typically have not taken into account the complexity of a daily industrial noise exposure. Further, in practice any concern for the temporal and level complexity of an industrial exposure is obviated by the acceptance of the equivalent energy principal as a guide to noise exposure criteria (e.g., ISO 1999). The equal energy hypothesis (EEH) initially proposed by Eldred *et al.* (1955), with early support from the guinea pig experiments of Eldredge and Covell (1958), suggests that equal amounts of acoustic energy (with some consideration given to frequency content) entering the ear will produce equal amounts of hearing loss under typical conditions of exposure. Over the ensuing years a number of studies have supported the concept of the EEH (Eldredge *et al.*, 1959; Dolan *et al.*, 1976; Clark, 1991; and others). This idea gave rise to the concept of the equivalent continuous noise level (L_{eq}), defined as the level of a continuous noise that, in the course of an 8 h workday, would cause the same sound energy to be received as that due to the actual noise over a typical workday. Energy as a basic metric for evaluating hazardous exposures was eventually extended to impact noise exposures (Atherley and Martin, 1971). While there is some recent support for an energy-based approach under limited exposure conditions (e.g., Qiu *et al.*, 2007), there is a consensus that such an approach will not cover all conditions of exposure (e.g., Lataye and Campo, 1996; Dunn *et al.*, 1991; Harding and Bohne, 2004; Hamernik *et al.*, 2003).

The impact that the temporal structure of an acoustic signal has on noise-induced hearing loss (NIHL) has been highlighted by the "toughening" and "conditioning" phenomena and an understanding of some of the cochlea's endogenous protective mechanisms underlying these phenomena. In the former case it has been shown (Miller *et al.*, 1963; Clark *et al.*, 1987; and others) that for subjects exposed to a daily interrupted noise, audiometric thresholds could improve despite the continuing exposure (i.e., the cochlea is gradually toughened). In the latter case, Canlon *et al.* (1988) and others have shown that exposure to a low level noise could reduce the effects of a subsequent traumatic exposure (i.e., the cochlea has been toughened by the low level conditioning noise). In a typical industrial noise exposure situation it could be argued that both of these phenomena would be operative. Temporal factors also become significant when the exposure is non-Gaussian (nonG), that is, when the noise contains transients that vary in their rate of appearance and in their distribution of peaks in the case of impacts, or their rms levels in the case of noise bursts. Such complex

---
[a)]Electronic mail: roger.hamernik@plattsburgh.edu

exposures were used by Hamernik *et al.* (2003) and Qiu *et al.* (2006) in showing the limitations of the energy metric and the advantage of using the statistical metric, kurtosis, in combination with energy to evaluate the hazard potential of a nonG exposure. They showed that for a constant energy there was an orderly relation between the kurtosis of an exposure and the resulting NIHL and sensory cell loss. As the kurtosis increased so did the hearing trauma. Experimental data on the effects of equivalent energy nonG exposures with time varying levels are not available.

In a recent study Qiu *et al.* (2007), using several different equivalent energy interrupted, intermittent, and time varying (IITV) Gaussian noise exposures, showed that despite some toughening effects the permanent threshold shifts (PTS) and sensory cell losses were reasonably similar across different IITV exposures and similar to an equivalent energy uniform and uninterrupted reference exposure. They concluded that the EEH is a reasonable approach to evaluating Gaussian noise exposures for hearing conservation purposes. The present paper, a continuation of the Qiu *et al.* (2007) work, presents data from a number of different nonG, IITV equivalent energy exposures. The exposures were designed to replicate some of the essential features of realistic industrial environments over a period of 3 weeks.

II. METHODS

A. Auditory evoked potential

One hundred and forty (140) chinchillas were used as subjects. Each animal was anesthetized [IM injection of Ketamine (35 mg/kg) and Xyalazine (1 mg/kg)] and made monaural by the surgical destruction of the left cochlea. During this procedure a bipolar electrode was implanted, under stereotaxic control, into the left inferior colliculus and the electrode plug cemented to the skull for the recording of auditory evoked potentials (Henderson *et al.*, 1973; Salvi *et al.*, 1982). The auditory evoked potential (AEP) was used to estimate pure tone thresholds. The animals were awake during testing and restrained in a yoke-like apparatus to maintain the animal's head in a constant position within the calibrated sound field. AEPs were collected to 20 ms tone bursts (5 ms rise/fall time) presented at a rate of 10/s. Each sampled wave form was analyzed for large-amplitude artifact, and, if present, the sample was rejected from the average and another sample taken. Averaged AEPs were obtained from 250 presentations of the 20 ms tone bursts. Thresholds were measured using an intensity series with 5 dB steps at octave intervals from 0.5 to 16 kHz. Threshold was defined to be one-half step size (2.5 dB) below the lowest intensity that showed a "response" consistent with the responses seen at higher intensities.

B. Experimental protocol

Following a 2 week postsurgical recovery, three AEP preexposure audiograms were obtained (on different days) on each animal at octave intervals between 0.5 and 16.0 kHz. If the mean of the three audiograms fell beyond one standard deviation (s.d.) of laboratory norms (Hamernik and Qiu, 2000), in the direction of poorer thresholds at more than one

test frequency, the animal was rejected. The animals were randomly assigned to one of 16 experimental groups with 7–16 animals/group.

The animals were exposed four (or less) at a time to one of the noise conditions detailed in the following. During exposure, animals were given free access to food and water and were rotated through a bank of six cages daily. The SPLs, across cages, in the middle of each cage, varied within less than ±1 dB. For the 5 day uniform, Gaussian or nonG continuous (uninterrupted) reference exposures, animals were removed daily for less than 0.5 h for AEP testing. A complete AEP audiogram was obtained on each animal each day of the exposure and the mean of the five audiograms defined asymptotic threshold (AT). For all the IITV groups the animals were tested at the end of the daily exposure on days 1, 17, 18, and 19. The difference between the threshold measured following the first day of exposure (T_1) and the mean of the thresholds measured following the last three days (T_{17-19}) of the exposure was accepted as an estimate of any threshold recovery or toughening (T_r) effect [i.e., $T_r = (T_1) - (T_{17-19})$]. Thirty days following the last exposure day for all exposure paradigms, three more audiograms were collected on different days and the mean used to define permanently shifted thresholds or permanent threshold shift.

C. Histology

Following the last AEP test protocol, each animal was euthanized under anesthesia and the right auditory bulla removed and opened to gain access to the cochlea for perfusion. Fixation solution consisting of 2.5% glutaraldehyde in veronal acetate buffer (final pH=7.3) was perfused through the cochlea. After 12–24 h of fixation the cochlea was postfixed in 1% OsO_4 in veronal acetate buffer. Surface preparation mounts of the entire organ of Corti were prepared (Engstrom *et al.*, 1966) and inner hair cell (IHC) and outer hair cell (OHC) populations were plotted as a function of frequency and location using the frequency-place map of Eldredge *et al.* (1981). Missing cells were identified by the presence of a characteristic phalangeal scar. For purposes of this presentation, sensory cell population data are presented as group averages (in percent missing) taken over octave band lengths of the cochlea centered on the primary AEP test frequencies.

D. Noise measurement and analyses

During the exposures the noise field was monitored with a Larson Davis 814 sound level meter equipped with a 1/2 in. microphone. The computer generated acoustic signal, produced by the Electro-Voice Xi-1152/94 speaker system, was transduced by a Brüel and Kjær 1/2 in. microphone (Model 4134), amplified by a Brüel and Kjær (Model 2610) measuring amplifier and fed to a Windows PC-based analysis system. The signal was sampled at 48 kHz with a recording duration of 5.5 min. The design and digital generation of the acoustic signal is detailed in Hsueh and Hamernik (1990, 1991). Calibration of the exposure was accomplished "off line" by analyzing two to eight, 5.5 min segments of the digitized acoustic wave form. In order to maintain equal en-
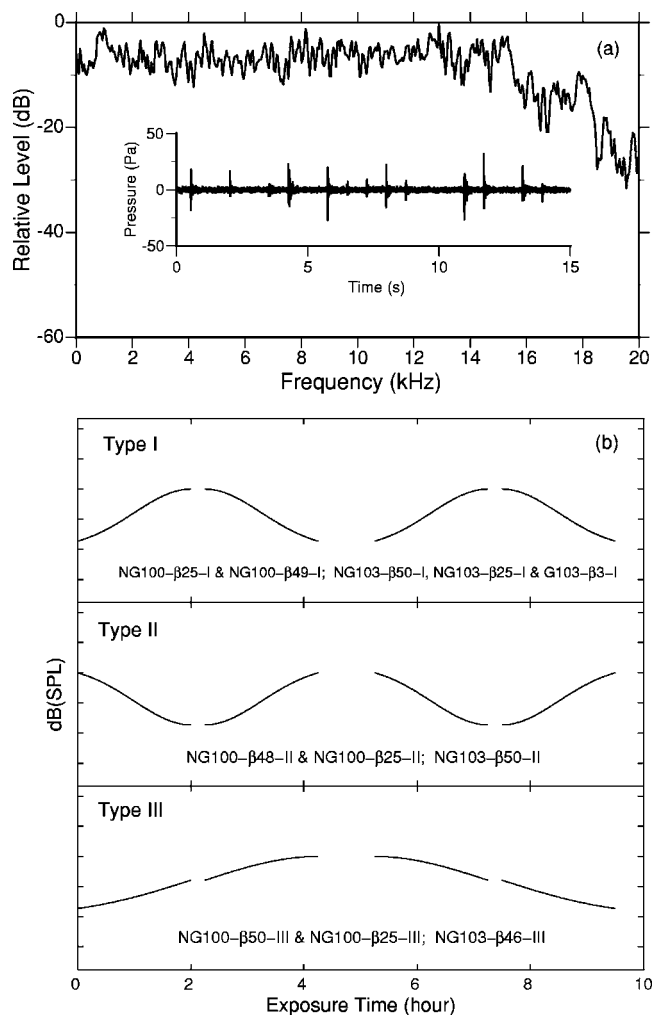
FIG. 1. (a) The relative spectral level of the unweighted 100 dB(A) SPL noise. The 103 dB(A) SPL noise had an unweighted spectrum that was qualitatively similar to the 100 dB(A) SPL noise. The inset shows a 15 s segment of a non-Gaussian noise. The impact peaks and interimpact intervals were randomly varied. (b) A sketch of the daily SPL variations that were used for the intermittent, interrupted, and time varying (IITV) noise exposure paradigm for the 100 and 103 dB(A) SPL groups. Groups exposed to each type of IITV exposure are identified. Each exposure sequence lasted for 19 days. Each daily exposure consisted of two 4.25 h periods with an hour break in between. Each 4.25 h exposure was interrupted for 15 min and each 5 day exposure sequence was separated by a 2 day break.

ergy exposure conditions across nonG noise exposed groups at the same overall SPL the peak and interval histograms of the impact transients and the level of the background noise (L_b), that is, the rms level of the noise between the transients, had to be adjusted. For the nonG, IITV exposures a mean value of the kurtosis, $\beta(t)$, was calculated by averaging the $\beta(t)$ from eight, 5.5 min samples of the exposure wave form.

E. Noise exposures

Hearing and cell loss data are presented from 16 different exposures. Each exposure had in common the same flat spectrum between 0.125 and approximately 15.0 kHz shown in Fig. 1(a). The nonG exposures were created by inserting impacts, having randomly varied peak levels and interimpact intervals, into a Gaussian background noise as described by Hsueh and Hamernik (1990, 1991). A 15 s sample of a nonG

wave form is shown in the inset of Fig. 1(a). The identification of each experimental group and the exposure parameters for the groups are detailed in Fig. 1(b) and Table I. Nine (9) exposures were presented at an overall SPL of 100 dB(A) and seven exposures were presented at 103 dB(A). All the 100 dB(A) exposures had the same total energy. Group G100-β3-U was exposed to a Gaussian noise and the remaining eight groups were exposed to a nonG noise that differed in either the value or range of values of the kurtosis or in the configuration of the SPL temporal variation [Fig. 1(b)]. Two of the nonG exposures (groups NG100-β25-U and NG100-β55-U) that were used as equal energy reference groups lasted for 24 h/day for 5 days and had a uniform, uninterrupted 100 dB(A) SPL. Group G100-β3-U, another reference group, was exposed to a uniform Gaussian noise for 5 days, 24 h/day. Six of the 100 dB(A) exposures were structured to model an idealized 3 week work shift with a varying noise level [IITV exposures, Fig. 1(b)]. Each daily IITV exposure consisted of two 4.25 h periods with an hour break in between. Each 4.25 h exposure was interrupted for 15 min and each 5 day sequence was separated by a 2 day break. Three different temporal variations in the sound pressure level (identified as type I, II, or III) were used for the exposures and are shown schematically in Fig. 1(b) along with the exposure group identification. It should be noted that there was no *a priori* reason for the choice of the Gaussian shaped patterns of sound level variation shown in Fig. 1(b).

There were seven groups exposed at an overall SPL of 103 dB(A). One of these groups (G103-β3-U) was exposed to a uniform Gaussian noise, 24 h/day for 5 days and one group (G103-β3-I) was exposed to a Gaussian, IITV noise with a type I SPL variation. One group (NG103-β50-U) was exposed to a uniform nonG noise, 24 h/day for 5 days. These three groups served as reference exposures for the five, nonG, IITV exposed groups. Summary AEP and sensory cell loss data across different exposure groups are compared in the figures identified in Table I.

F. Statistical analysis

Threshold shifts and the percent sensory cell loss in octave-band lengths of the cochlea were compared among the groups of animals for each noise exposure level [100 and 103 dB(A) SPL] using a mixed model analysis of variance (ANOVA) with repeated measures on one factor (frequency). The probability of a type I error was set at 0.05. Statistically significant main effects of frequency were expected and found in all of the following analyses because of the frequency-specific nature of the audibility curve of the chinchilla and the noise exposure stimulus. For this reason the main effects of frequency are not addressed in the following presentation of the results.

III. RESULTS AND DISCUSSION

A. Preexposure thresholds

The group mean preexposure AEP thresholds are shown as symbols in Fig. 2. The solid and broken lines show the mean thresholds for all the animals exposed at 100 and

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Hamernik *et al.*: Noise trauma and equal energy hypothesis   2247

TABLE I. Outline of the noise exposure conditions for all experimental groups.

Fig.	Group I.D.	Group size	Noise type[a]	Exposure type	L_{eq} dB(A)	L_{eq} range dB(A)	Mean kurtosis	Kurtosis range	Peak SPL (dB)	L_b dB(A)[b]
3	G103-β3-U	N=8	G	Uniform 5 day	103	⋯	3	⋯	⋯	⋯
4 and 5	NG103-β50-U	N=7	nonG	Uniform 5 day	103	⋯	50	[43, 57]	[122, 131]	93
	NG103-β46-III	N=8	nonG	IITV 19 day, III	103	[89, 106]	46	[35, 55]	[104, 134]	[80, 97]
	NG103-β50-I	N=8	nonG	IITV 19 day, I	103	[89, 106]	50	[33, 56]	[105, 134]	[81, 98]
	NG103-β50-II	N=7	nonG	IITV 19 day, II	103	[89, 106]	50	[38, 62]	[102, 134]	[80, 97]
6 and 7	G103-β3-I	N=8	G	IITV 19 day, I	103	[89, 105]	3	⋯	⋯	⋯
	NG103-β25-I	N=7	nonG	IITV 19 day, I	103	[89, 105]	25	[19, 32]	[101, 132]	[81, 98]
3	G100-β3-U	N=16	G	Uniform 5 day	100	⋯	3	⋯	⋯	⋯
8	NG100-β25-U	N=12	nonG	Uniform 5 day	100	⋯	25	[23, 28]	[115, 129]	95
	NG100-β25-III	N=8	nonG	IITV 19 day, III	100	[67, 105]	25	[17, 31]	[80, 130]	[60, 98]
	NG100-β25-II	N=7	nonG	IITV 19 day, II	100	[69, 105]	25	[18, 35]	[90, 130]	[63, 97]
	NG100-β25-I	N=8	nonG	IITV 19 day, I	100	[70, 105]	25	[16, 35]	[92, 130]	[62, 98]
9	NG100-β55-U	N=12	nonG	Uniform 5 day	100	⋯	55	[46, 62]	[121, 132]	0
	NG100-β50-III	N=8	nonG	IITV 19 day, III	100	[68, 106]	50	[33, 68]	[84, 132]	[60, 97]
	NG100-β48-II	N=8	nonG	IITV 19 day, II	100	[70, 105]	48	[35, 57]	[91, 132]	[63, 97]
	NG100-β49-I	N=8	nonG	IITV 19 day, I	100	[70, 105]	49	[31, 57]	[92, 133]	[63, 97]

[a]G=Gaussian; nonG=non-Gaussian.
[b]L_b, the level of the Gaussian background noise.

103 dB(A) respectively. Standard errors were small and all animals and groups fell within ±1 s.d. of laboratory norms based on 1572 chinchillas (shaded area). These data are consistent with published thresholds for the chinchilla (Fay 1988). There were no systematic differences in preexposure thresholds among the experimental groups.

B. The 100 and 103 dB(A) Gaussian reference exposures

Figure 3 presents the results of exposure to a uniform Gaussian noise at 100 and 103 dB(A) for 5 days, 24 h/day (groups G100-β3-U and G103-β3-U, respectively). The exposure was interrupted for only approximately 20 min each

FIG. 2. Mean preexposure thresholds (symbols) for the 16 experimental groups defined in Table I. The shaded area represents the mean preexposure thresholds ±1 s.d. of the laboratory norm based on 1572 chinchillas. The solid line represents the mean preexposure thresholds of the 87 chinchillas included in the groups exposed to the 100 dB(A) SPL noises and the broken line the preexposure thresholds for the 53 chinchillas exposed at 103 dB(A).

day for AEP testing in order to obtain estimates of asymptotic threshold shift (ATS). The upper panels show the AEP data and the lower panels the inner and outer hair cell loss data. The shaded area in the upper panels represents the PTS. The difference between AT and the preexposure thresholds is an estimate of the ATS. The bars on the data symbols in this and subsequent figures represent the standard error (s.e.) of the mean. If a bar is not present the s.e. is less than the size of the symbol.

These two exposures, at each level, serve as reference points for all the Gaussian and nonG, IITV or uninterrupted exposure groups discussed in the following. For the 100 dB(A) exposure (group G100-β3-U) ATS increased with increasing frequency from about 37 dB at 0.5 kHz to 73 dB at 16 kHz, while PTS varied from near 0 dB at 0.5 kHz to 24 dB at 16 kHz. The IHC loss did not exceed 10% at the higher frequencies and OHC losses varied from less than 10% to almost 20% in the highest frequency band. Increasing the level to 103 dB(A) (group G103-β3-U) produced from about 65 dB ATS at 16 kHz to a maximum of 84 dB at 4 kHz. During the 5 day exposure there were large shifts in threshold across the entire range of AEP test frequencies. PTS increased with increasing frequency from 10 to 30 dB. IHC loss was small with about a 12% loss at 8 kHz and no loss at and below 2 kHz. OHC loss showed a bimodal distribution with 26% losses in the 0.250 and 0.5 kHz bands and 36% loss at 8 kHz with little loss in the 1.0 and 2.0 kHz region. It should be noted that any thresholds at 90 dB seen in Fig. 3 or in any of the subsequent figures represent the upper limit of the AEP test system and a threshold shift computed at such frequencies represents an estimate of the lower bound only. The most noticeable effect of the 3 dB increase in level is the increased threshold shifts
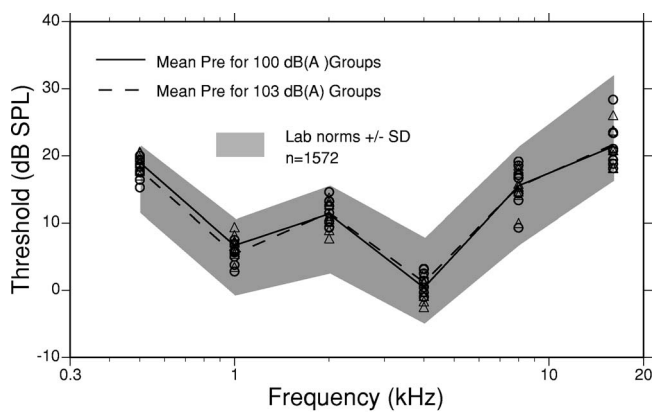
Hamernik *et al.*: Noise trauma and equal energy hypothesis

FIG. 3. Summary data for reference group G100-β3-U exposed to the 100 dB(A) SPL Gaussian, uninterrupted 5 day exposures and group G103-β3-U exposed to a similar noise but at 103 dB(A) SPL. Upper panels show the group means preexposure thresholds, asymptotic threshold (AT), and 30 day postexposure thresholds for each group. The lower panels show the group mean percent outer and inner sensory cell (OHC and IHC) losses for each group. Error bars indicate the standard error (s.e.) of the mean. If no error bar is present the s.e. is smaller than the size of the symbol.

and OHC loss especially at the low frequencies. A two-way ANOVA confirmed these differences. There was a statistically significant main effect of group for the ATS, PTS, and OHC loss and no significant main effect of group for IHC loss. There was no significant interaction of group and frequency for PTS but there was a significant interaction for ATS, IHC loss, and OHC loss.

C. The 103 dB(A) exposures

All the exposures discussed in the following had the same total energy and the same spectrum. They differ only in the nature of their level variations, their kurtosis, and whether or not the exposure was interrupted and intermittent.

1. The effects of exposures having an equal energy and kurtosis but a variable level

Figure 4 presents complete group mean data sets for four of the 103 dB(A), nonG exposure conditions. Figure 4 illustrates the type of data set that was obtained for each experimental group and from which any subsequent summary figures are distilled. The effects of three different nonG, IITV exposures (groups NG103-β46-III, NG103-β50-I, and NG103-β50-II) and an uninterrupted nonG, 5 day reference exposure (NG103-β50-U) are shown. The four exposures had approximately the same mean value of the kurtosis [$\beta(t) \sim 50$] and the same total energy and spectra. For group NG103-β50-U exposed for 5 days to the uninterrupted nonG noise, group mean asymptotic thresholds were 90 dB across the range of test frequencies. These threshold values do not represent actual AEP thresholds but rather the upper limit of the AEP measurement system. After a 30 day recovery period, thresholds recovered to between 48 and 80 dB SPL with the higher frequencies showing the highest thresholds. Group mean thresholds for the three IITV exposures measured following the first day's exposure were similarly at or very near the limit of the AEP test system except for the 0.5 and 1.0 kHz test frequencies in group NG103-β46-III where thresholds of about 76 dB SPL were measured. All three

IITV exposures produced a T_r that was greatest at the lower frequencies and disappeared at the highest test frequencies. While a T_r was measured, the true magnitude could not be determined at a number of frequencies as a result of the ceiling effect in the AEP test system. All the IITV exposures produced a T_r and when thresholds were not limited by the ceiling effect, T_r was as much as 22 dB at 0.5 kHz for the 103 dB(A) exposures. During the 30 day recovery period, thresholds only recovered 10–25 dB leaving each group with as much as 65 dB permanent losses. Inner and outer hair cell losses generally reflect the severity of the shifted thresholds. Outer hair cell loss is severe and in some frequency regions complete, with greatest loss occurring at the higher frequencies. Inner hair cell losses are also severe with maximum loss in the 4 kHz region of the basilar membrane. Threshold and cell loss data from these four groups are summarized and compared in Fig. 5. Apparent from Fig. 5 is the similarity of the PTS, IHC loss, and OHC loss across the four groups. A two-way ANOVA confirmed that there was no significant main effect of group and no interaction between group and frequency for PTS, IHC loss, and OHC loss.

Based on the above presented results, nonG exposures of equivalent energy, spectra, and kurtosis can produce equivalent trauma whether or not the exposure is interrupted, intermittent, or follows a different temporal pattern of level variation. This result is similar to the conclusions drawn in a recent paper (Qiu et al., 2007) where Gaussian IITV exposures were used. The toughening, T_r, seen in the IITV exposures did not seem to have any effect on the final PTS. While an accurate T_r cannot be measured because of the AEP ceiling effect, all the IITV exposures at 103 dB(A) did produce toughening but no subsequent protective effect. Similar levels of PTS were found in both the uninterrupted exposures where there was no toughening and the IITV exposures that showed some toughening. Also note the considerable increase in trauma associated with these nonG exposures when compared to the equivalent energy Gaussian [$\beta(t)=3$] reference exposure shown in Figs. 3(c) and 3(d). The increase in trauma from nonG exposures has been documented from a
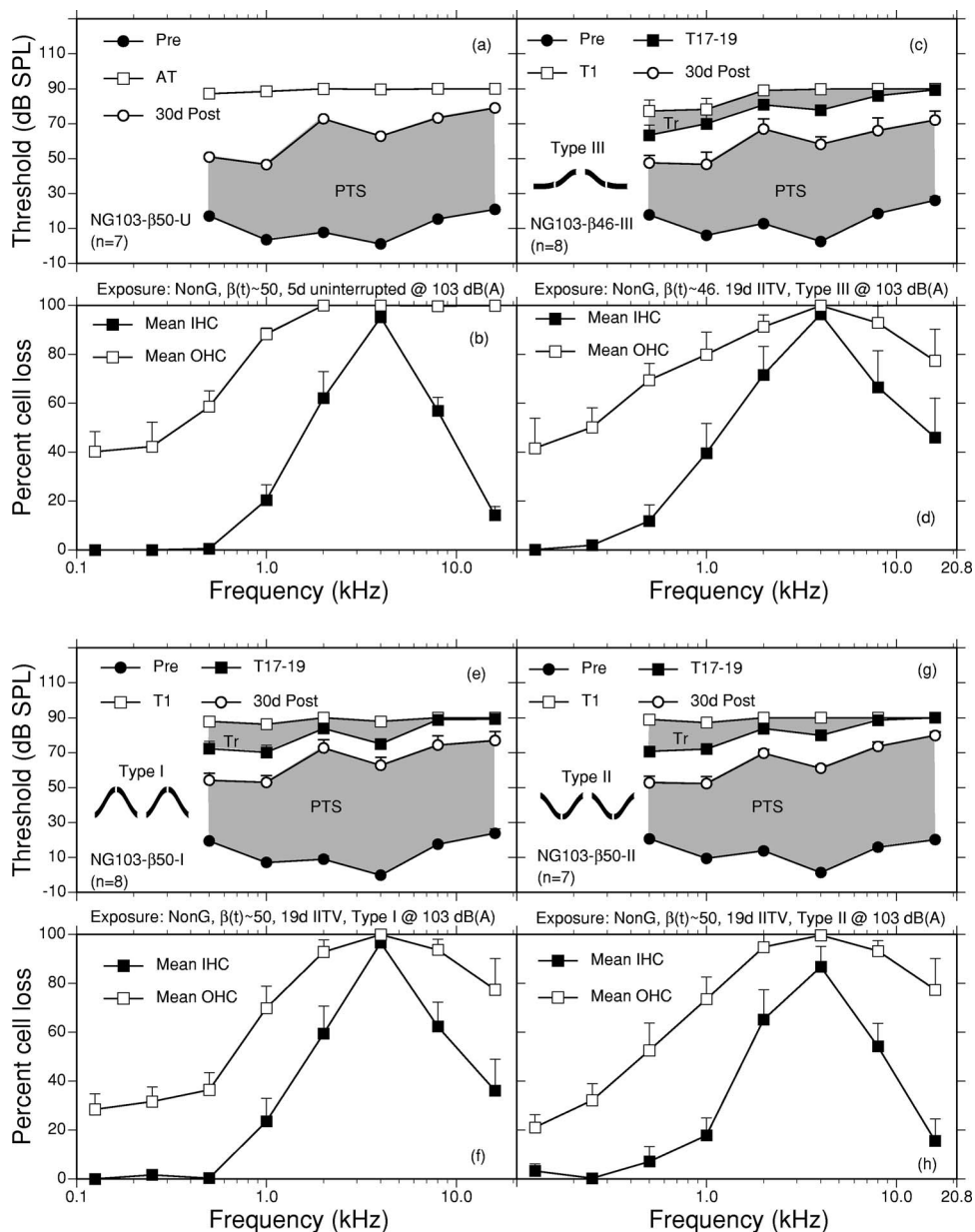
FIG. 4. Summary data from four of the groups exposed to 103 dB(A) SPL, non-G, $\beta(t) \sim 50$, 5 or 19 day noise paradigms. (a) The group (NG103-β50-U) mean preexposure thresholds, AT, and 30 day postexposure thresholds for the 5 day uninterrupted, uniform reference exposure. (b) The corresponding group mean outer and inner sensory cell (OHC, IHC) losses. (c), (d); (e), (f); and (g), (h) A similar presentation of data for the three IITV equivalent energy exposures, NG103-β46-III, NG103-β50-I, and NG103-β50-II, respectively. Instead of AT, the group mean thresholds measured immediately following exposure on day 1 (T_1) and the group mean thresholds measured immediately following exposure on days 17, 18, and 19 (T_{17-19}) are shown. Toughening (T_r) and permanent threshold shifts (PTS) are shown shaded. Error bars indicate the s.e. of the mean. If no error bar is present the s.e. is smaller than the size of the symbol.

variety of exposures that incorporated impacts or noise bursts to produce nonG exposure conditions (Hamernik *et al.*, 2003) as well as from comparisons of equal energy continuous noises and pure impact noise exposures (Dunn *et al.*, 1991).

2. The effects of IITV exposures having an equivalent energy and temporal structure but varying in their kurtosis

Figure 6 presents the group mean data from groups G103-β3-I and NG103-β25-I. Animals in group G103-β3-I were exposed to a 19 day, IITV Gaussian noise, $\beta(t)=3$ and group NG103-β25-I to a 19 day, IITV nonG noise, $\beta(t) \sim 25$. Group NG103-β50-I discussed earlier [Figs. 4(e) and 4(f)] was exposed to the same noise except with a $\beta(t) \sim 50$. All three exposures had a type I level variation. All three exposures produced large increases in threshold (T_1) following the first day of exposure that varied from about 68 to the 90 dB SPL limit of the AEP test system. The nonG

groups showed the greatest increases in threshold. All three exposures produced similar amounts of toughening T_r with a maximum T_r of about 22 dB at the lower frequencies and very little or no toughening at the highest frequencies. This was similar in magnitude and frequency specificity to the exposures discussed above in Sec. III C 1. Following a 30 day recovery period, thresholds had recovered $22-46$ dB in the group exposed to the Gaussian IITV noise (G103-β3-I), while thresholds in the nonG noise exposed group NG103-β25-I only recovered $16-26$ dB. Inner and outer hair cell loss was also considerably less in the Gaussian noise exposed group G103-β3-I. A summary and comparison of the data from the three 103 dB(A) exposures with the type I level variation but having different $\beta(t)$ is shown in Fig. 7. Figures 7(a)–7(c) show the PTS, %IHC, and %OHC loss for the three groups, respectively. There is a clear ordering of PTS, IHC loss, and OHC loss. The Gaussian noise exposed group G103-β3-I showed a 10 dB PTS at the lowest frequencies that increased to 45 dB at the highest test fre-

Hamernik *et al.*: Noise trauma and equal energy hypothesis
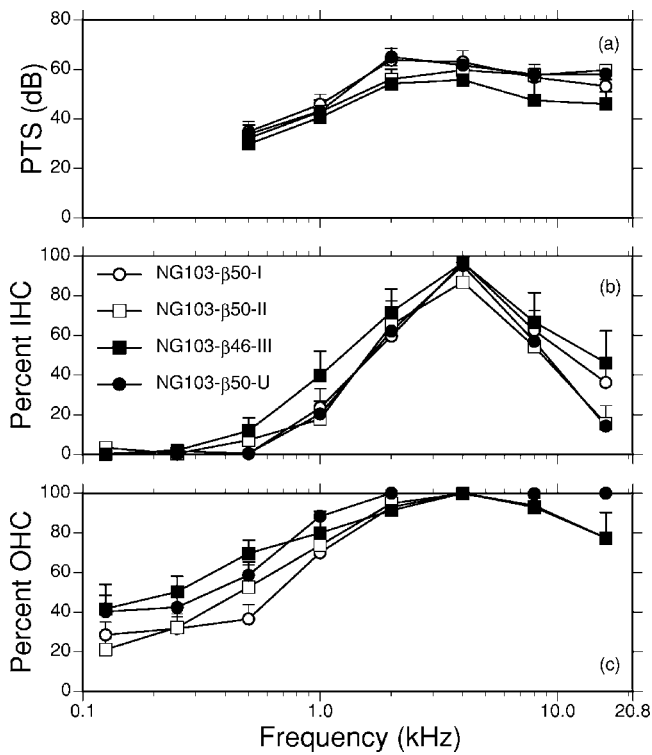
FIG. 5. A comparison of the PTS and cell loss data for the four groups shown in Fig. 4. (a) The PTS. (b) The percent IHC loss. (c) The percent OHC loss. All four groups were exposed to a nonG noise having the same energy and kurtosis. For group NG103-β50-U the uniform exposure lasted for 5 uninterrupted days. Groups NG103-β50-II, NG103-β46-III, and NG103-β50-I received the 19 day, IITV exposures. Error bars indicate the s.e. of the mean. If no error bar is present the s.e. is smaller than the size of the symbol.
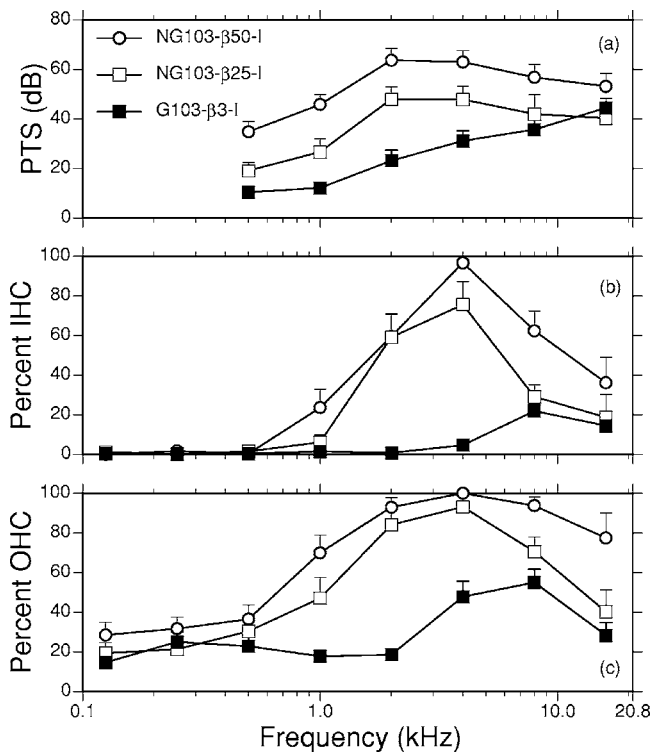
FIG. 7. A comparison of the PTS and cell loss data for the two groups shown in Fig. 6 along with the data from group NG103-β50-I (Fig. 4). (a) The PTS. (b) The percent IHC loss. (c) The percent OHC loss. All three groups were exposed to an IITV noise having the same energy and a type I level variation but different mean $\beta(t)$. For group G103-β3-I, $\beta(t) = 3$ (i.e., a Gaussian noise); for group NG103-β25-I, $\beta(t) \sim 25$; and for group NG103-β50-I, $\beta(t) \sim 50$. Error bars indicate the s.e. of the mean. If no error bar is present the s.e. is smaller than the size of the symbol.

quency. Group NG103-β25-I, the nonG, $\beta(t) \sim 25$ exposure was intermediate in loss with 19–47 dB PTS that peaked at 2 kHz. The greatest loss was seen in the high kurtosis, $\beta(t) \sim 50$, nonG noise exposed group NG103-β50-I. PTS in this group varied from 35 dB at 0.5 kHz to a maximum of 64 dB at 2.0 kHz. IHC and OHC losses generally followed the trends seen in the PTS with maximum IHC losses of 75%

and 95% in the 4 kHz region of the basilar membrane for the $\beta(t) \sim 25$ and $\beta(t) \sim 50$ exposures, respectively. The pattern of loss for the nonG exposures was broadly distributed across frequency. The Gaussian exposure produced maximum IHC loss of 22% in the 8 kHz region and little or no loss at the lower frequencies. OHC losses were considerably larger; more broadly distributed and generally paralleled the
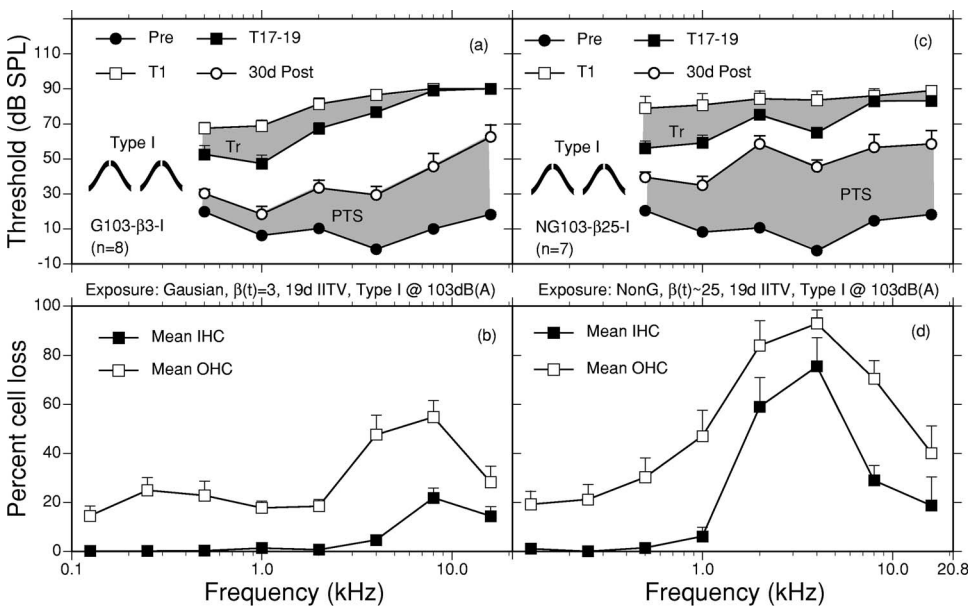
FIG. 6. Summary data from two of the groups exposed at 103 dB(A) SPL, to a 19 day, type I, IITV noise paradigm. Group G103-β3-I was exposed to a Gaussian $\beta(t) = 3$ noise while group NG103-β25-I was exposed to a non-G, $\beta(t) \sim 25$ noise. (a) The group (G103-β3-I) mean thresholds measured immediately following exposure on day 1 (T_1) and the group mean thresholds measured immediately following exposure on days 17, 18, and 19 (T_{17-19}). T_r and PTS are shown shaded. The group mean preexposure thresholds, and 30 day postexposure thresholds are also shown. (b) The group mean outer and inner sensory cell (OHC, IHC) losses. (c), (d) A similar presentation of data for group NG103-β25-I. Error bars indicate the s.e. of the mean. If no error bar is present the s.e. is smaller than the size of the symbol.
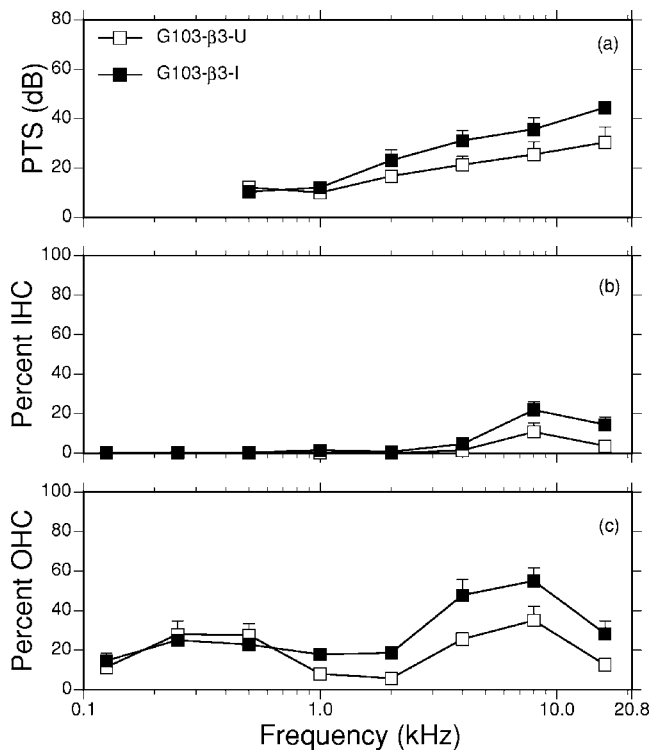
FIG. 8. A comparison of the (a) PTS, (b) IHC loss, and (c) OHC loss measured in group G103-β3-U exposed to a 5 day, uniform uninterrupted Gaussian noise and group G103-β3-I exposed to a 19 day Gaussian type I, IITV noise (Fig. 6) of equivalent energy. Error bars indicate the s.e. of the mean. If no error bar is present the s.e. is smaller than the size of the symbol.

FIG. 9. A comparison of the (a) PTS, (b) IHC loss, and (c) OHC loss measured in four groups exposed at 100 dB(A) SPL noise at $\beta(t) \sim 25$. The three IITV groups NG100-β25-III, NG100-β25-I, and NG100-β25-II were exposed to a 19 day noise with type I, II and III level variations respectively. Group NG100-β25-U received an uninterrupted, uniform 5 day exposure. Error bars indicate the standard error (s.e.) of the mean. If no error bar is present the s.e. is smaller than the size of the symbol.

IHC loss. A two-way ANOVA confirmed these differences. For PTS, IHC loss, and OHC loss there was a statistically significant main effect of group and a significant interaction of group and frequency.

In this set of IITV exposures, differing only in the value of the kurtosis, there is a clear difference in trauma produced as a result of increasing the kurtosis; trauma increases with an increase in $\beta(t)$. These results are in agreement with the results of Hamernik et al. (2003) where a very different exposure paradigm was followed. Results from a variety of exposure conditions are now available showing that the kurtosis of the amplitude distribution of a noise is a variable that should be considered when trying to establish the hazards to hearing posed by long-term exposure to acoustically diverse noise exposures.

3. A comparison between the 5 day uninterrupted and 19 day IITV Gaussian exposures

Based on the results of Qiu et al. (2007) where Gaussian uninterrupted and IITV exposures at 100 and 106 dB(A) SPL were used one would anticipate that groups G103-β3-U [Figs. 3(c) and 3(d)] and G103-β3-I [Figs. 6(a) and 6(b)] would produce similar levels of trauma. These two groups are compared in Fig. 8. An ANOVA analysis of these data indicates that there is a significant main effect of group for the PTS and no interaction between group and frequency. For the IHC and OHC loss there is a significant main effect of group and an interaction of group and frequency. These differences are relatively small, amounting to a 12 dB maxi-

mum difference at 16 kHz in the PTS, 12% difference in the IHC loss at only 8 and 16 kHz, and less than 22% for OHC loss at the higher frequencies. The more traumatic of the two exposures was the 19 day IITV exposure. This is contrary to what would be expected since the IITV exposure produced a modest toughening effect and the interrupted paradigm should have given ample opportunity for thresholds to recover between the daily and weekly exposures.

D. The 100 dB(A) exposures

As discussed earlier (Fig. 3), the 100 dB(A) Gaussian reference exposure produced statistically significant less ATS, PTS, and OHC loss than did the 103 dB(A) Gaussian exposure. The results from the 103 dB(A) nonG exposures (Figs. 4 and 5) suggested that the equal energy principle applied to IITV exposures provided that the exposures had the same $\beta(t)$. In an effort to confirm this result, several groups of animals were exposed at 100 dB(A) SPL to various nonG, IITV exposures at two values of $\beta(t)$.

In addition to the IITV groups, one group at each $\beta(t)$ received an uninterrupted nonG exposure for 5 day, 24 h/day. Data summaries for the four groups (NG100-β25-U, NG100-β25-I, NG100-β25-II, NG100-β25-III) exposed to a $\beta(t) \sim 25$ noise are shown in Fig. 9 and for the $\beta(t) \sim 50$ groups (NG100-β55-U, NG100-β49-I, NG100-
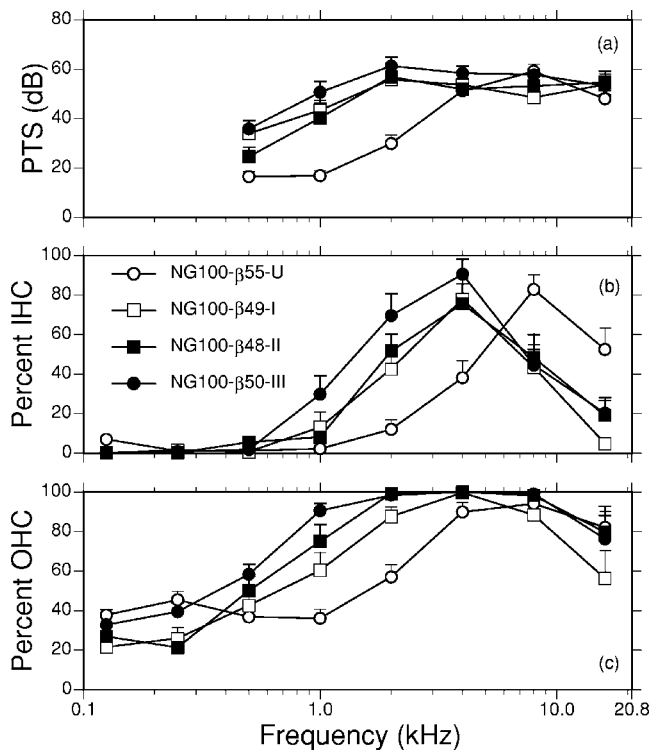
FIG. 10. A comparison of the (a) permanent threshold shift (PTS), (b) inner hair cell loss (IHC) and (c) outer hair cell loss (OHC) measured in four groups exposed at 100 dB(A) SPL noise at $\beta(t) \sim 50$. The three IITV groups NG100-β50-III, NG100-β49-I and NG100-β48-II were exposed to a 19 day noise with type I, II and III level variations respectively. Group NG100-β55-U received an uninterrupted, uniform 5 day exposure. Error bars indicate the standard error (s.e.) of the mean. If no error bar is present the s.e. is smaller than the size of the symbol.

β48-II, NG100-β50-III) in Fig. 10. Groups NG100-β25-U with $\beta(t) \sim 25$ and NG100-β55-U with $\beta(t) \sim 50$ received the 5 day, uninterrupted nonG exposure.

The two-way ANOVA on the data in Fig. 9 indicated that there was no significant main effect of group for the PTS, IHC loss, and OHC loss but there was a significant interaction of group and frequency. This interaction, indicating frequency specific effects, is the result of a shift in the pattern of PTS and sensory cell loss toward the higher frequencies for group NG100-β25-U, the group exposed for 5 continuous days, relative to the three IITV groups (NG100-β25-I, NG100-β25-II, NG100-β25-III). When an ANOVA is performed on only the three IITV groups, there is no interaction effect and no main effect of group for the OHC or IHC loss. There is, however, a main effect of group for the PTS. Group NG100-β25-III with the type III level variation tended to have the greater PTS, although the differences overall were small except at 2 kHz where the PTS exceeded that of the other two IITV groups by about 18 dB.

Increasing the kurtosis of four similar exposures from $\beta(t) \sim 25$ to $\beta(t) \sim 50$ produced an increased level of trauma (Fig. 10) as anticipated but an almost parallel set of frequency specific results to those shown in Fig. 9 for the $\beta(t) \sim 25$ exposures. A two-way ANOVA on the data in Fig. 10 indicated a significant interaction of group and frequency for all dependent variables as well as a main effect of group for the PTS and OHC loss but no main effect of group for the

IHC loss. An ANOVA on the data from only the three groups receiving the IITV exposures (i.e., eliminating group NG100-β55-U that received the 5 day exposure) indicated no main effect of group and no interaction of group and frequency for all dependent variables. Thus all three IITV exposures at the higher level of kurtosis produced the same degree of trauma regardless of the type of level variation.

The data from the four exposures shown in Fig. 5 were consistent in showing the same effects on hearing regardless of whether the exposures were 19 day, IITV, or 5 day continuous, thus supporting the use of an energy metric for evaluating equivalent $\beta(t)$ exposures. Also surprising was that the 5 day uninterrupted exposure at 103 dB(A) produced the same effects as the three IITV exposures. Lowering the overall level of the noise by 3 dB did show that the IITV exposures were statistically similar regardless of the type of level variation as long as $\beta(t)$ was the same. However, unlike the results at 103 dB(A) shown in Fig. 5, there were small but significant differences between the 5 day uninterrupted and 19 day IITV exposures.

The toughening effects (T_r) measured from exposure to the 100 dB(A), IITV exposures were similar in magnitude and frequency distribution to that reported earlier for the 103 dB(A) exposures with over 20 dB of toughening being measured at 0.5 and 1.0 kHz. Since asymptotic thresholds at many frequencies were not limited by the AEP ceiling effect more accurate estimates of T_r could be made. As with the 103 dB, IITV groups this amount of toughening did not reduce the trauma. In general, the 100 dB(A), IITV exposures produced more PTS and cell loss at most frequencies than did the uninterrupted 5 day exposures (Figs. 9 and 10). A number of papers have shown that, as one might expect, interrupted exposures will typically produce less trauma than equivalent energy continuous exposures (Patuzzi, 1998; Campo and Lataye, 1992; and others). In the above-noted exposure paradigms the IITV exposures that produced modest levels of toughening also produced the same or more trauma than the uninterrupted reference exposures at the same $\beta(t)$.

This surprising result was also seen in the Gaussian IITV exposures reported by Qiu *et al.* (2007) and in the long-term impact noise exposures reported by Hamernik and Ahroon (1998).

IV. CONCLUSION

The above-presented data along with the results in Qiu *et al.* (2007) show that for a given total energy, temporal variations in the rms level of a noise exposure do not have a significant effect on the degree of trauma associated with Gaussian and nonG complex noise exposures. The data are also consistent in showing that under a variety of exposure conditions the kurtosis of the amplitude distribution is an exposure variable that should be considered when evaluating complex noise exposures for the purpose of hearing conservation practice. Increasing the kurtosis of an exposure at the same energy increases the PTS and sensory cell loss while exposures at the same kurtosis and energy regardless of their temporal complexity produce similar levels of trauma.

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Hamernik *et al.*: Noise trauma and equal energy hypothesis   2253

These data as well as the data from a number of referenced papers showing both support for and against the EEH suggest that while energy is a necessary metric for the evaluation of noise environments for hearing conservation purposes it is not sufficient. However, the results presented here along with those of Qiu *et al.* (2006, 2007) and Hamernik *et al.* (2003) suggest that energy and kurtosis may represent a necessary and sufficient set of metrics for such an evaluation. A better understanding of the role of the kurtosis metric in NIHL should lead to its incorporation into a new generation of more predictive damage risk criteria for noise exposure provided that human exposure and hearing loss data can be acquired from suitably designed epidemiological studies.

Atherley, G. R. C., and Martin, A. M. (**1971**). "Equivalent-continuous noise level as a measure of injury from impact and impulse noise," Ann. Occup. Hyg. **14**, 11–28.

Campo, P., and Lataye, R. R. (**1992**). "Intermittent noise and equal energy hypothesis," in *Noise Induced Hearing Loss*, edited by A. Dancer, D. Henderson, R. J. Salvi, and R. P. Hamernik (Mosby Year Book, St. Louis), pp. 456–466.

Canlon, B., Borg, E., and Flock, A. (**1988**). "Protection against noise trauma by pre-exposure to a low-level acoustic stimulus," Hear. Res. **34**, 197–200.

Clark, W. W. (**1991**). "Recent studies of temporary threshold shift (TTS) and permanent threshold shift (PTS) in animals," J. Acoust. Soc. Am. **90**, 155–163.

Clark, W. W., Bohne, B. A., and Boettcher, F. A. (**1987**). "Effects of periodic rest on hearing loss and cochlear damage following exposure to noise," J. Acoust. Soc. Am. **82**, 1253–1264.

Dolan, T. R., Murphy, R. J., and Ades, H. W. (**1976**). "A comparison of the permanent deleterious effects of intense noise on the chinchilla resulting from either continuous or intermittent exposure," in *Effects of Noise on Hearing*, edited by Henderson, D. Hamernik, R. P. Dosanjh, and D. S. Mills (Raven, New York), pp. 327–340.

Dunn, D. E., Davis, R. R., Merry, C. J., and Franks, J. R. (**1991**). "Hearing loss in the chinchilla from impact and continuous noise exposure," J. Acoust. Soc. Am. **90**, 1979–1985.

Eldred, K. M., Gannon, W. J., and Von Gierke, H. E. (**1955**). "Criteria for short time exposureof personnel to high intensity jet aircraft noise," U. S. Air Force, WADC Technical Note 55–355, Wright-Patterson Air Force Base, OH.

Eldredge, D. H., and Covell, W. P. (**1958**). "A laboratory method for the study of acoustic trauma," Laryngoscope **68**, 465–477.

Eldredge, D. H., Covell, W. P., and Gannon, R. P. (**1959**). "Acosutic trauma following intermittent exposure to tones," Ann. Otol. Rhinol. Laryngol. **68**, 723–733.

Eldredge, D. H., Miller, J. D., and Bohne, B. A. (**1981**). "A frequency-position map for the chinchilla cochlea," J. Acoust. Soc. Am. **69**, 1091–1095.

Engstrom, H., Ades, H. W., and Andersson, A. (**1966**). *Structural Pattern of the Organ of Corti* (Almqvist and Wiksell, Stockholm, Sweden).

Fay, R. R. (**1988**). *Hearing in Vertebrates* (Hill-Fay Associates, Winnetka, IL).

Hamernik, R. P., and Ahroon, W. A. (**1998**). "Interrupted noise exposures: Threshold shift dynamics and permanent effects," J. Acoust. Soc. Am. **103**, 3478–3488.

Hamernik, R. P., and Qiu, W. (**2000**). "Correlation among evoked potential thresholds, otoacoustic emissions and cochlear sensory cell loss following various noise exposures in the chinchilla," Hear. Res. **150**, 245–257.

Hamernik, R. P., Qiu, W., and Davis, B. (**2003**). "The effects of the amplitude distribution of equal energy exposures on noise-induced hearing loss: The kurtosis metric," J. Acoust. Soc. Am. **114**, 386–395.

Harding, G. W., and Bohne, B. A. (**2004**). "Noise-induced haircell loss and total exposure energy: Analysis of a large data set," J. Acoust. Soc. Am. **115**, 2207–2220.

Henderson, D., Hamernik, R. P., Woodford, C., Sitler, R. W., and Salvi, R. J. (**1973**). "Evoked response audibility curve of the chinchilla," J. Acoust. Soc. Am. **54**, 1099–1101.

Hsueh, K. D., and Hamernik, R. P. (**1990**). "A generalized approach to random noise synthesis: Theory and computer simulation," J. Acoust. Soc. Am. **87**, 1207–1217.

Hsueh, K. D., and Hamernik, R. P. (**1991**). "Performance characteristics of a phase domain approach to random noise synthesis," Noise Control Eng. J., **36**, 18–32.

International Organization for Standardization, (ISO–1999). (**1990**). "Acoustics: Determination of occupational noise exposure and estimation of noise-induced hearing impairment," ISO-1999, Geneva.

Lataye, R., and Campo, P. (**1996**). "Applicability of the Leq as a damage-risk criterion: An animal experiment," J. Acoust. Soc. Am. **99**, 1621–1632.

Miller, J. D., Watson, C. S., and Covell, W. P. (**1963**). "Deafening effects of noise on the cat," Acta Suppl. **176**, 1–91.

Patuzzi, R. (**1998**). "Exponential onset and recovery of temporary threshold shift after loud sound: Evidence for long term inactivation of mechano-electrical transduction channels," Hear. Res. **125**, 17–38.

Qiu, W., Davis, B., and Hamernik, R. P. (**2007**). "Hearing loss from interrupted, intermittent, and time varying Gaussian noise exposures: The applicability of the equal energy hypothesis," J. Acoust. Soc. Am. **121**, 1613–1620.

Qiu, W., Hamernik, R. P., and Davis, B. (**2006**). "The kurtosis metric as an adjunct to energy in the prediction of trauma from continuous, nonGaussian noise exposures," J. Acoust. Soc. Am. **120**, 3901–3906.

Salvi, R. J., Ahroon, W. A., Perry, J., Gunnarson, A., and Henderson, D. (**1982**). "Psychophysical and evoked-response tuning curves in the chinchilla," Am. J. Otolaryngol. **3**, 408–416.

Taylor, W., Lempert, B., Pelmear, P., Hemstock, I., and Kershaw, J. (**1984**). "Noise levels and hearing thresholds in the drop forging industry," J. Acoust. Soc. Am. **76**, 807–819.

Phantom echo highlight amplitude and temporal difference resolutions of an echolocating dolphin, *Tursiops truncatus*

Mark W. Muller[a)]
Marine Mammal Research Program, Hawaii Institute of Marine Biology, 46-007 Lilipuna Road, Kaneohe, Hawaii 96744-3617 and Department of Mechanical Engineering, University of Hawaii at Manoa, 2540 Dole Street, Honolulu, Hawaii 96822-2303

Whitlow W. L. Au and Paul E. Nachtigall
Marine Mammal Research Program, Hawaii Institute of Marine Biology, 46-007 Lilipuna Road, Kaneohe, Hawaii 96744-3617

John S. Allen III
Department of Mechanical Engineering, University of Hawaii at Manoa, 2540 Dole Street, Honolulu, Hawaii 96822-2303

Marlee Breese
Marine Mammal Research Program, Hawaii Institute of Marine Biology, 46-007 Lilipuna Road, Kaneohe, Hawaii 96744-3617

A dolphin's ability to discriminate targets may depend greatly on the relative amplitudes and the time separations of echo highlights within the received signal. Previous experiments with dolphins varied the physical parameters of targets, but did not fully investigate how changes in these parameters correspond with the scattered acoustic wave forms and the dolphin's subsequent response. This experiment utilizes a phantom echo system to test a dolphin's detection response to relative amplitude differences of secondary and trailing echo highlights and the time separation differences of all the echo highlights both within and outside the animal's integration window. By electronically manipulating the amplitude and temporal separation of the echo highlights, the underlying acoustic classification cues are more efficiently investigated. The animal successfully discriminated between a standard echo signal and one with the secondary highlight amplitude lowered by 7 dB from the standard. Furthermore, the animal successfully discriminated between a standard echo signal and one with the trailing highlight amplitude lowered by 3 dB from the standard and also a standard echo signal and one with a time separation of 150 μs between the secondary and trailing highlights. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2769973]

I. INTRODUCTION

One of the outstanding characteristics of the bottlenose dolphin sonar system which distinguishes it from manmade sonar systems is its ability to make fine distinctions in the features of target signals. Previously, a study was conducted using an echolocating dolphin and physical targets that consisted of a hollow aluminum cylinder for the standard target. Hollow aluminum cylinders of the same outer diameter and length as the standard target but with various wall thicknesses were used for the comparison targets (Au and Pawloski, 1992). It was discovered that the animal could discriminate between the standard target and a comparison target varying ±0.3 mm in the wall thickness. Other biosonar experiments have been conducted to investigate the capabilities of dolphins to discriminate between objects differing in structure, shape, and material composition (see Nachtigall,

1980, and Au, 1998, for a review). While targets with varying physical differences (such as materials, wall thickness, etc.) offer empirical verification of a dolphin's abilities, they do not always provide distinct and clear acoustic cues for further experimental analysis and the development of corresponding models and algorithms.

Important acoustic classification cues are more efficiently examined by electronically manipulating the echoes returned to the animal (Aubauer *et al.*, 2004). A study on dolphin discrimination of complex (multi-component) synthetic echoes in noise has been investigated and it was found that a higher-amplitude leading echo component did not interfere with discrimination of a lower-amplitude trailing echo component (Helweg *et al.*, 2003), i.e., the dolphin could successfully discriminate between differences in secondary echo components that followed larger-amplitude primary echo components. This suggests that the subject in Helweg *et al.* (2003) may have processed multiple echo components as separable features. However, the synthetic echoes used in that study did not simulate an echo signal from a physical

---
a)Author to whom correspondence should be addressed. Electronic mail: mwmuller@gmail.com

target. An important consideration in these biosonar experiments is to determine which acoustic features change corresponding to the changes of the physical characteristics of the targets. Although the outgoing echolocation signals are brief (40–75 μs), the echoes reflected from objects can be hundreds of microseconds in duration and contain a rich structure that encodes information about the target's characteristics (e.g., Chapman, 1971; Gaunaurd et al., 1998; Urick, 1983). The diversity of complex time and frequency-domain structures includes great variability in the amplitude ratio of multiple echo components, referred to as the highlights. The variance in echo structures between targets immediately raises questions about how dolphins exploit the complex timing and relative amplitude of highlight structure to detect and discriminate targets.

While the integration time of the dolphin was measured to be 264 μs (Au et al., 1988), this does not necessarily mean that the dolphin is not sensitive to highlight patterns within the integration time. Various discrimination experiments, like the wall-thickness discrimination (Au and Pawloski, 1992), indicate the dolphin may be sensitive to highlight differences—both time separation differences and amplitude differences within the integration window. This implies that the dolphin may be utilizing variations in an echo within the integration window. The following study was conducted to determine the natural capability of an echolocating *Tursiops truncatus* to discriminate phantom echoes with differences in highlight amplitudes and time separations between highlights. Phantom echoes have been used in previous experiments (Aubauer and Au, 1998; Aubauer et al., 2000) and the dolphin responds to phantom echoes as though the signals are from physical targets. The use of phantom echoes allows fine experimental control over the amplitude, timing, and spectral characteristics of the echo highlights within the phantom echo signal associated with a stainless-steel spherical target with a diameter of 76.2 mm (3 in.).

II. METHODS

A. Subject

The subject was BJ, a 21-year-old female Atlantic bottlenose dolphin housed in an open water floating pen complex at the Hawaii Institute of Marine Biology in Kaneohe Bay. This dolphin has over 16 years of experience as a psychoacoustical research subject and was also the subject in the phantom echo studies of Aubauer and Au (1998) and Aubauer et al. (2000). Based on auditory evoked potential measurements conducted in 2001 and 2005 (Fig. 1), her high-frequency hearing threshold is relatively low compared to typical dolphin hearing thresholds ($\sim$100 kHz) (Johnson, 1967). BJ is trained on a Go/No-Go paradigm (Schusterman, 1980).

B. Phantom echo system

A phantom echo system was fabricated to detect and record outgoing echolocation clicks and transmit a single phantom echo per detected click. The constructed system was similar to that described by Au et al. (1987), Aubauer and Au (1998), and Aubauer et al. (2000). However the sys-
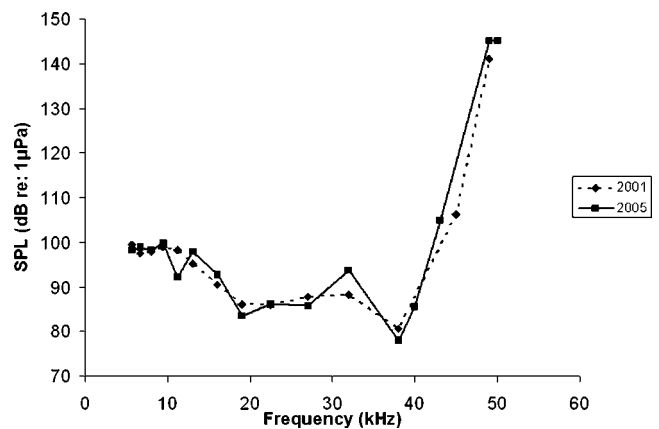
FIG. 1. Auditory evoked potential measurements conducted on BJ in 2001 and 2005 showing auditory thresholds as a function of frequency. BJ's hearing rapidly decreases after 45 kHz.

tem developed for this study has higher processing speeds than the previously published systems which allow it to effectively operate up to 100 clicks/s without missing any clicks. The system also incorporates custom written LABVIEW software to generate phantom echo signals as opposed to the Phantom Echo Generator used in the previously published systems. Figure 2 displays a schematic of the phantom echo setup. The dolphin produces a click which is first detected by a Brüel and Kjaer 8103 triggering hydrophone with a flat frequency response ($\pm$3 dB) up to 120 kHz. The signal triggers the system to begin recording with a second Brüel and Kjaer 8103 hydrophone. The two hydrophones are 5 cm apart and the time delay between the signal arriving at the first hydrophone and the second hydrophone allows the system to be triggered and to record the entire click with the second hydrophone. The recorded click is amplified and sent to a Measurement Computing Corporation PCI-DAS4020/12 analog-to-digital input board which digitizes the signal at a sampling rate of 1 MHz. The signal is then processed by a custom written LABVIEW program which convolves the sig-
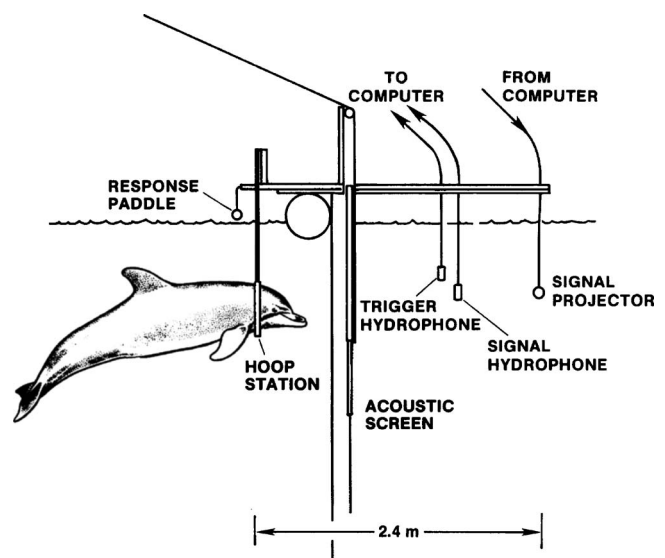
FIG. 2. Experimental configuration with echolocating dolphin and phantom echo system (after Au et al.1987).
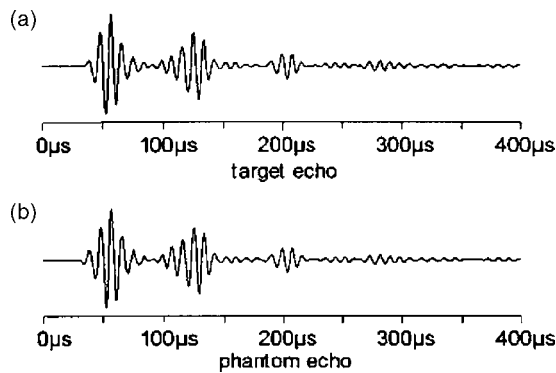
FIG. 3. Comparison of (a) real and (b) phantom echoes of a solid stainless-steel sphere (after Aubauer and Au, 1998).

nal with the transfer function for a stainless-steel spherical target.

The convolution allows the system to produce phantom echoes that accurately reflect physical target echoes in terms of response to intensity and frequency content changes of the clicks (Fig. 3). The convolution method can be used to transform any dolphin echolocation signal into phantom echoes of high fidelity and complexity (Aubauer *et al.*, 2000). Given an incident dolphin echolocation signal $s_t(t)$, the backscattered target echo $e_t(t)$ can be found by convolution of $s_t(t)$ and the target impulse response $h(t)$. The target impulse response can be used to describe the acoustic backscattering processes of targets (Aubauer and Au, 1998). The echo can be expressed in both the time and the frequency domain

$$e_t(t) = h(t) * s_t(t) = \int_{-\infty}^{\infty} h(\tau) \cdot s_t(t - \tau)d\tau, \tag{1}$$

$$E_t(\omega) = H(\omega) \cdot S_t(\omega), \tag{2}$$

where the target transfer function $H(\omega)$ and the spectra $E_t(\omega)$, $S_t(\omega)$ are the Fourier transforms of the target impulse response, the echo, and the source click signal, respectively.

The resulting signal is convolved with a transducer equalization function and sent to a Strategic Test UF6000 digital to analog wave form generator board which outputted the signal to an International Transducer Corporation 1042 transducer at an update rate of 1 MHz. The LABVIEW program also coordinated the transfer of information between the input analog-to-digital board and the output digital-to-analog board. The resulting echo is then propagated back to the dolphin.
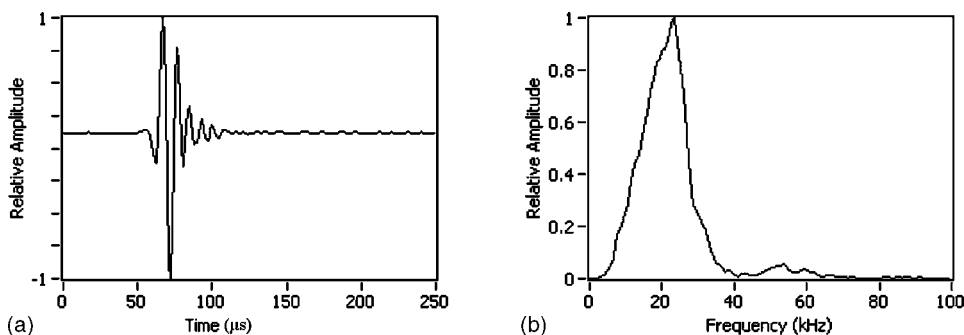
C. Behavioral paradigm

The data collection sessions began with BJ facing the trainer, touching her rostrum against a foam stationing pad located at the water surface. Upon presentation of a hand cue, the dolphin would submerge, swim 10 m across the pen, and position her head in the test station hoop (see Fig. 2). The trainer then raised the acoustically opaque stainless-steel screen, which cued the dolphin to freely echolocate and also cued the experimenter to activate the phantom echo system. The screen remained raised for a 3 s trial, during which time BJ received a phantom echo for each click she produced. A correct response when the standard stimulus was presented was made if BJ swam out of the hoop and touched a nearby foam paddle with her rostrum within 3 s after the acoustic screen was lowered. A correct response when a comparison stimulus was presented was made if BJ remained in the hoop for 3 s after the acoustic screen was lowered. Both correct responses were reinforced by a bridging stimulus (the trainer's whistle) and a consistent fish reward. Data were collected using a modified Gellerman series (Gellerman, 1933) that had been counterbalanced in ten-trial blocks. Each session was initiated with a ten-trial block of warm-up trials. One session was run per day; up to five sessions were run per week over a period of four months.

D. Phantom echo stimuli

Figure 4 displays the time domain wave form and the frequency spectrum of a typical echolocation signal recorded from BJ during phantom echolocation trials. Figures 5(a) and 5(b) display the standard stainless-steel sphere phantom echo signal and spectrum which are calculated from (1) and (2), respectively, with the outgoing click shown in Fig. 4. The phantom echo signal contains multiple components, or highlights. For the purposes of this experiment, the phantom echo signal can be divided into three highlights (Fig. 5(c)). The first highlight (i) is the primary highlight and is on the order of $50-75$ μs in duration, depending on the incident echolocation signal. The second highlight (ii) is the secondary highlight and is on the order of $75-100$ μs in duration. The third highlight group (iii) is the trailing highlights and is on the order of $150-200$ μs in duration.

There were four conditions used for the comparison trials. The first variable manipulated was the amplitude of the secondary highlight which was varied between -12 and -6 dB from the standard. Figure 6(a) displays an example of a comparison stimulus used for the first condition a stainless-

FIG. 4. A typical echolocation click and the corresponding spectrum recorded from BJ during a phantom echolocation experiment.

Muller *et al.*: Dolphin discrimination of echo highlight differences   2257
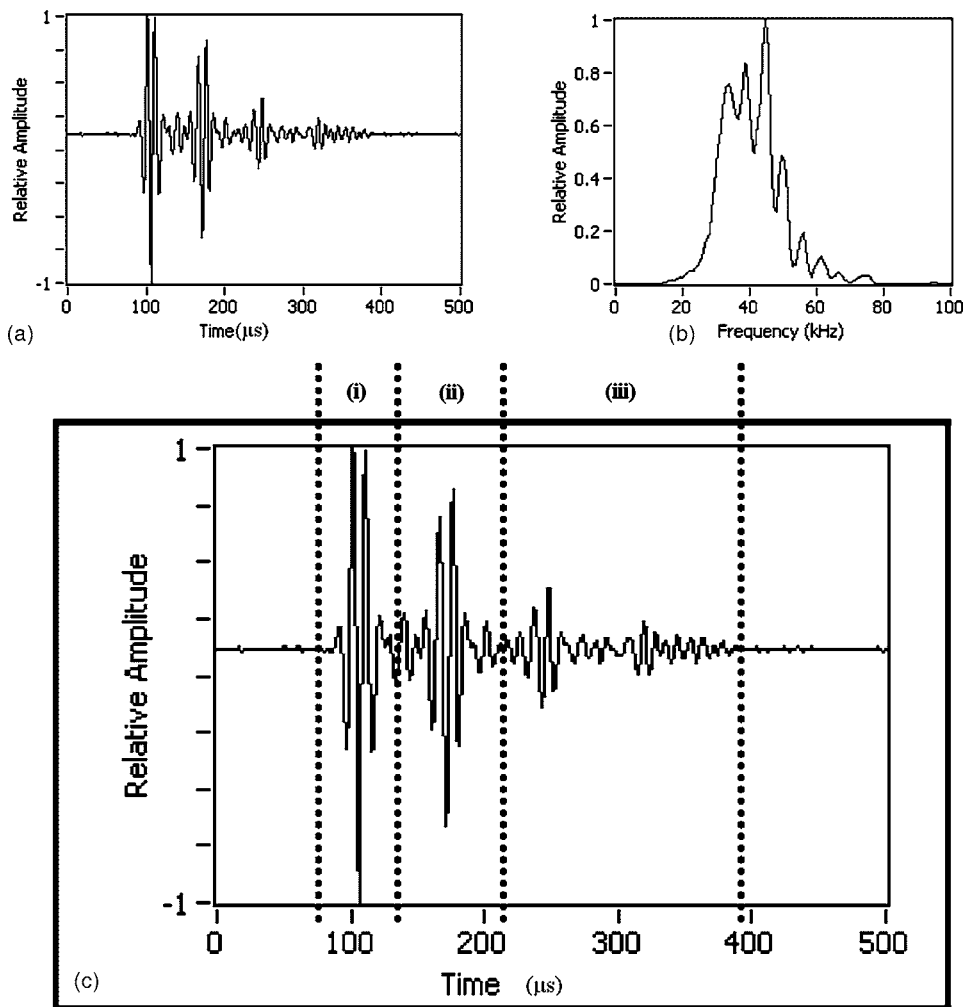
FIG. 5. (a), (b) Standard stainless-steel sphere phantom echo signal and corresponding spectrum measured underwater from the echolocation click displayed in Fig. 4; (c) standard stainless-steel sphere phantom echo with highlights displayed as (i) primary, (ii) secondary, and (iii) trailing.

steel sphere phantom echo with the amplitude of the secondary highlight lowered by 6 dB from the standard. The frequency spectra of Fig. 5(b) and Fig. 6(b) are similar; however, it can be determined from the wave forms that a considerable amount of energy has been eliminated from the secondary highlight.

The next variable manipulated was the amplitude of the trailing highlights, which was varied between −12 and −1 dB from the standard. Figure 6(c) displays an example of a comparison stimulus used for the second condition; a stainless-steel sphere phantom echo with the amplitude of the trailing highlight lowered by 6 dB from the standard. There is a greater difference between the frequency spectra of Fig. 6(d) and the standard than between Fig. 6(b) and the standard. By examining the wave form it can be determined that most of the energy of the trailing highlight has been eliminated.

The third variable manipulated was the temporal separation between the first and second highlights, which was initially varied by 500 μs from the standard. The 500 μs temporal separation was initially chosen because it was the largest value that could be used in the customized LABVIEW program. Figure 6(e) displays an example of a comparison stimulus used for the third condition; the temporal separation between the primary and secondary highlights increased by 125 μs from the standard. The fourth variable manipulated was the temporal separation between the secondary and trail-

ing highlights, which was varied between 100 and 500 μs from the standard. Figure 6(g) displays an example of a comparison stimulus used for the fourth condition; the temporal separation between the secondary and trailing highlights increased by 125 μs from the standard. There are significant differences between the spectra of the comparison stimuli with the temporal separations between highlights (Figs. 6(f) and 6(h)) and the spectrum of the standard (Fig. 5(b)).

E. Experimental design

A modified method of constants (Green and Swets, 1966) similar to the method followed by Helweg *et al.* (2003) in their study of dolphin discrimination of complex synthetic echoes was used. First, the amplitude of the secondary highlight was manipulated while holding the amplitude of the primary and trailing highlights constant. A running estimate of percent correct was calculated for each session for all four conditions using a ten-block sliding window and the 75%-correct point(s) were tabulated (Helweg *et al.*, 2003). A set of five secondary highlight amplitude values was tested per session. The dolphin's performance was measured as percent correct for each secondary highlight amplitude condition. Next the amplitude of the trailing highlight was manipulated while holding the amplitude of the primary and secondary highlights constant. A set of five trailing high-
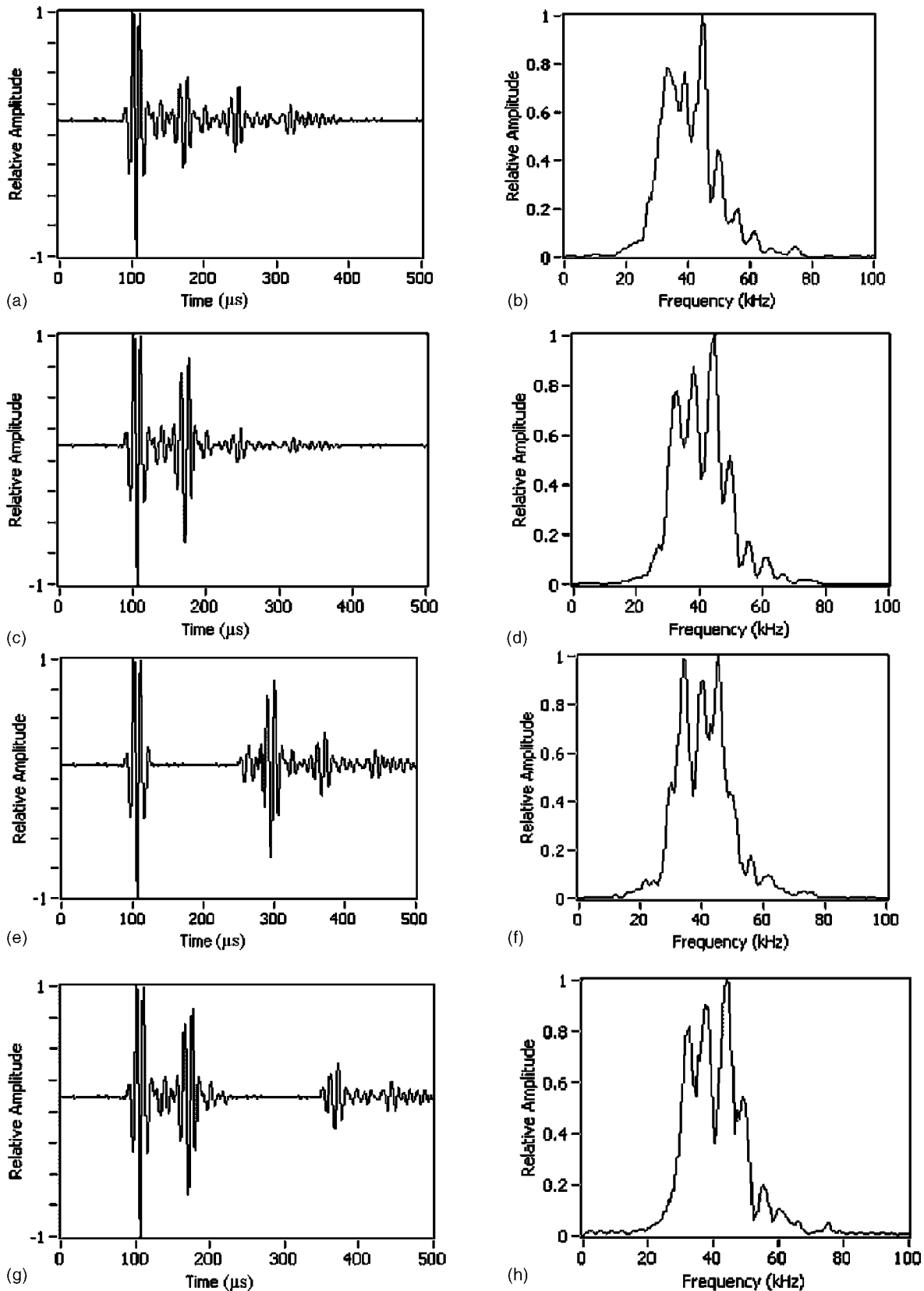
FIG. 6. (a) Stainless-steel sphere phantom echo with second highlight lowered by 6 dB from the standard and (b) corresponding spectrum; (c) stainless-steel sphere phantom echo with trailing highlights lowered by 6 dB from the standard and (d) corresponding spectrum; (e) stainless-steel sphere phantom echo with temporal separation between the first and second increased by 125 μs from the standard and (f) corresponding spectrum; (g) stainless-steel sphere phantom echo with temporal separation between the second and trailing highlights increased by 125 μs from the standard and (h) corresponding spectrum.

light amplitude values was tested per session. The dolphin's performance was measured as percent correct for each trailing highlight amplitude condition for the trailing highlights amplitude.

The same method was used for the highlight temporal separation experiments. First, the temporal separation between the primary and secondary highlights was manipulated while holding the temporal separation between the secondary and trailing highlights constant. A set of five temporal separation values between the primary and secondary highlights was tested per session. The dolphin's performance was measured as percent correct for each temporal separation condition between the primary and secondary highlights. Next, the temporal separation between the secondary and trailing highlights was manipulated while holding the temporal separation between the primary and secondary highlights constant. A set of five temporal separation values between the secondary and trailing highlights was tested per session. The dolphin's performance was measured as percent correct for each temporal separation condition between the secondary and trailing highlights.

The values for the amplitude difference and temporal separation difference conditions were chosen using a modified up-down procedure similar to the method used by Helweg *et al.* (2003). The highlight amplitude manipulation experiments began with a −12 dB difference in highlight amplitude between the standard and comparison stimuli. Upon successful performance, the highlight amplitude was changed to a −6 dB difference between the standard and the comparison. Depending on performance, the highlight amplitude was changed by ±3 dB from the previous amplitude setting. As a final step, depending on the animal's performance, the highlight amplitude was changed in steps of ±1 dB from the previous amplitude setting until a 1 dB discrimination threshold was reached. The temporal difference experiments began with a 500 μs difference in highlight temporal separation between the standard and the comparison stimuli. Upon successful performance, the highlight temporal separation was changed to a 250 μs difference between the standard and the comparison stimuli. Depending on performance, the highlight temporal separation was changed by ±125 μs from the previous temporal difference setting. As a final step, depending on the animal's performance, the highlight temporal separation was changed in steps of ±25 μs from the previous temporal difference setting until a 25 μs discrimination threshold was reached.

III. RESULTS

The dolphin's performance for the secondary highlight amplitude manipulation experiment is shown in Fig. 7(a), with percent correct discrimination plotted as a function of the highlight amplitude difference for discrimination. The dolphin could no longer successfully discriminate between the standard phantom echo and the comparison phantom echo with the manipulated secondary highlight when the secondary highlight amplitude was −6 dB from the standard. The dolphin's performance for the trailing highlights amplitude manipulation experiment is shown in Fig. 7(b), with
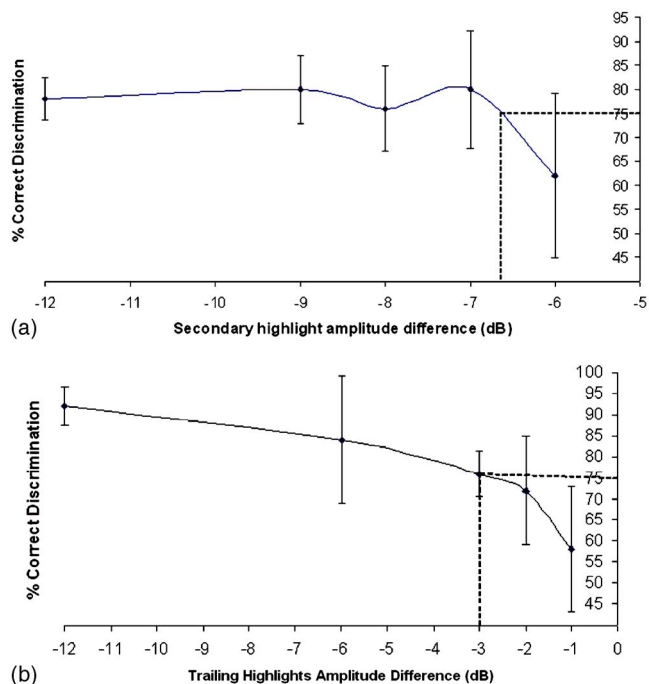
FIG. 7. (Color online) (a) Dolphin's performance results as a function of the secondary amplitude difference between the comparison and standard stimuli are shown; (b) dolphin's performance results as a function of the trailing amplitude difference between the comparison and standard stimuli are shown. Each data point represents the result of 50 trials.

percent correct discrimination plotted as a function of the amplitude difference for discrimination. The dolphin could no longer successfully discriminate between the standard phantom echo and the comparison phantom echo with the manipulated trailing highlights when the trailing highlights amplitude was −2 dB from the standard.

The animal did not reach a successful discrimination performance when presented with the phantom echo with the 500 μs temporal separation between the primary and secondary highlights after 100 trials. The dolphin responded correctly for 58% of the comparison trials where the comparison stimuli were presented. The dolphin successfully discriminated between the standard and comparison echoes at a 72% correct discrimination level. The dolphin's performance for the temporal separation between the secondary and trailing highlights experiment is shown in Fig. 8, with percent correct discrimination plotted as a function of the temporal difference. The dolphin could no longer successfully discriminate between the standard phantom echo and the comparison phantom echo with the manipulated temporal separation between the secondary and trailing highlights when temporal separation was 125 μs from the standard.

IV. DISCUSSION

The animal successfully discriminated between a standard phantom echo signal and one with the secondary highlight amplitude lowered by 7 dB from the standard. Also a successful discrimination was observed for a standard phantom echo signal and one with the trailing highlight amplitude lowered by 3 dB from the standard. Moreover, the animal successfully discriminated between a standard phantom echo

Muller *et al.*: Dolphin discrimination of echo highlight differences
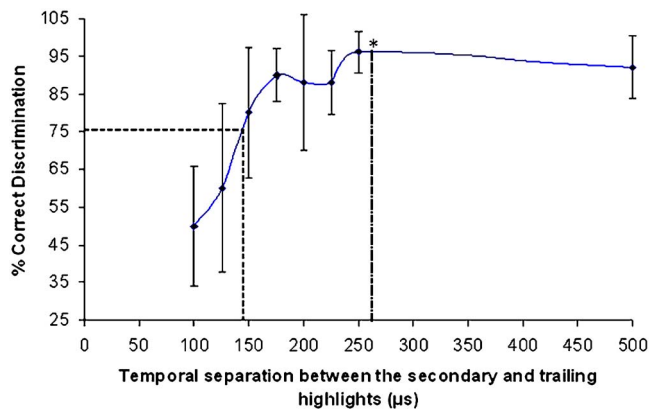
FIG. 8. (Color online) Dolphin's performance results as a function of the temporal difference between the secondary and trailing highlights between the comparison and standard stimuli are shown. Each data point represents the result of 50 trials. The asterisk represents the point of the dolphin's auditory integration time (264 μs).

signal and one with a time separation of 150 μs between the secondary and trailing highlights. The animal did not successfully discriminate between a standard phantom echo signal and one with a 500 μs temporal separation between the primary and secondary highlights.

The study of Au *et al.* (1988) reports that for short duration, broadband, high-frequency sonar signals, the auditory system of a dolphin behaves like an energy detector with an integration time of 264 μs. Examination of the energy density of the phantom echo signals, as given

$$E = 10 \log \left[\int_0^\infty p^2(t) dt \right], \tag{3}$$

where $p(t)$ is the instantaneous acoustic pressure of the phantom echo signal, reveals the energy density difference between the standard and the comparison with a −6 dB secondary highlight modification is approximately twice the energy density difference between the standard and the comparison with the −3 dB trailing highlight modification. The duration of the trailing echo highlights is also approximately twice the duration of the second echo highlight. This result suggests that the dolphin's energy detection resolution of highlights may improve as the length of the highlight signal is increased, i.e., the longer the duration of the highlight signal, the easier it may be to detect energy differences in individual echo highlights. This is consistent with the findings Plomp and Bouman (1959) and Zwislocki (1960) in that typically there is a 3 dB improvement in discrimination threshold for each doubling of the signal duration.

It is not clear as to why the animal did not successfully discriminate the comparison phantom echo with the 500 μs temporal difference between the first and second highlights from the standard. This experiment was performed directly following the conclusion of the amplitude manipulation experiments. Only the second and trailing highlights were altered in the amplitude manipulation experiment so it is possible that the dolphin learned to ignore the first highlight and pay closer attention to the secondary and trailing highlights. Helweg *et al.* (2003) suggested that the dolphin auditory system may be able to reject higher-amplitude highlights while attending to lower-amplitude echo highlights related to objects of interest. If the primary highlight is ignored, the standard and comparison phantom echoes are exactly identical for any temporal separation length between the primary and secondary echo highlights. Another possible reason that the animal did not discriminate successfully is that it may have not understood the task. For example, a confused dolphin might have perceived the comparison phantom echo with the 500 μs temporal separation as two independent signals since the 500 μs temporal separation is greater than its integration time. These explanations may potentially explain the poor discrimination performance of the animal during the first temporal separation experiment.

For the temporal separation of the secondary and trailing highlights experiment, the dolphin's performance was high for time separations greater than 125 μs. Once the temporal separation falls to 125 μs or less, the performance drops to below the successful discrimination performance criterion. However, it is interesting to note that there was not a direct drop off. For time separations just inside the integration time, there is a slight drop in performance followed by an increase in performance before the final drop off. There is also high variability in this region and future work should address any temporal window just inside and outside the integration time to determine how the dolphin performs.

For targets which vary in material composition or density, the scattering resonances will be different thus resulting in amplitude and frequency differences in the secondary highlights. Since a dolphin can discriminate between target signals varying slightly in amplitude of a single highlight, the highlight amplitude difference may be a cue used by the animal in order to discriminate between targets varying only in composition or density. Targets that vary in material composition will have different longitudinal and transverse sound velocities such that these different sound speeds will affect the time arrival of the individual echo highlights. The dolphin can discriminate between the target signals slightly varying in temporal separation between highlights. Therefore, the temporal separation of the highlights may be a cue used by the animal in order to discriminate between targets which vary only in composition. However, when a broadband auditory stimulus has multiple, correlated components with a temporal delay between the components, such as the stimuli used in these experiments, it is possible that a time separation pitch (with a frequency equivalent to the inverse of the temporal delay) will be perceived. Au and Hammer (1980) and Hammer and Au (1980) suggested that dolphins may use time separation pitch cues to discriminate targets. Therefore, it cannot be ruled out that the subject in this study may have been using time separation pitch cues when the discriminations of the phantom echo stimuli were performed.

Based on the results, a dolphin can isolate and process brief acoustic features that lie within and between energy integration windows of the phantom echoes. The energy detection mechanism may be more sensitive to acoustic features of longer durations within the integration window. The animal may be able to reject higher amplitude phantom echo features when performing discrimination tasks in order to focus on lower amplitude highlights. The dolphin may be

performing an energy summation within the integration time while looking for amplitude and temporal variations of an echo within the integration time. However, the details of the interplay between the integration time and echo highlight discrimination of dolphins are not well understood. Further investigation is needed in this respect to better understand a dolphin's echolocation discrimination abilities.

ACKNOWLEDGMENTS

Au, W. W. L. (**1998**). "Sonar target detection and recognition by odontocetes," in *Animal sonar: Processes and Performance*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 451–465.

Au, W. W. L. and Hammer, C. E. (**1980**). "Target recognition via echolocation by Tursiops truncatus," in *Animal sonar systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 855–858.

Au, W. W. L., Moore, P. W. B., and Martin, S. W. (**1987**). "Phantom electronic target for dolphin sonar research," J. Acoust. Soc. Am. **82**(2), 711–712.

Au, W. W. L., Moore, P. W. B., and Pawloski, D. A. (**1988**). "Detection of complex echoes in noise by an echolocating dolphin," J. Acoust. Soc. Am. **83**(2), 662–668.

Au, W. W. L., and Pawloski, D. A. (**1992**). "Cylinder wall thickness difference discrimination by an echolocating Atlantic bottlenose dolphin," J. Comp. Physiol., A **170**, 41–47.

Aubauer, R., and Au, W. W. L. (**1998**). "Phantom echo generation: A new technique for investigating dolphin echolocation," J. Acoust. Soc. Am. **104**(3), 1165–1170.

Aubauer, R., Au, W. W. L., and Nachtigall, P. E. (**2004**). "Acoustic simulation of phantom target echoes in dolphin behavioral experiments," in *Echolocation in Bats and Dolphins*, edited by J. A. Thomas, C. F. Moss, and M. Vater (University of Chicago Press, Chicago), pp. 514–518.

Aubauer, R., Au, W. W. L., Nachtigall, P. E., Pawloski, D. A., and DeLong, C. M. (**2000**). "Classification of electronically generated phantom targets by an Atlantic bottlenose dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **107**(5), 2750–2754.

Chapman, S. (**1971**). "Size, shape, and orientation of sonar targets measured remotely," Am. J. Phys. **39**, 1181–1190.

Gaunaurd, G. C., Brill, D., Huang, H., Moore, P. W. B., and Strifors, H. C. (**1998**). "Signal processing of the echo signatures returned by submerged shells insonified by dolphin "clicks." Active classification," J. Acoust. Soc. Am. **103**, 1547–1557.

Gellerman, L. W. (**1933**). "Chance orders of alternative stimuli in visual discrimination experiments," J. Genet. Psychol. **42**, 206–208.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Hammer, C. E. andAu, W. W. L., (**1980**). "Porpoise echo-recognition: An analysis of controlling target characteristics," J. Acoust. Soc. Am. **68**(5), 1285–1293.

Helweg, D. A., Moore, P. W., Dankiewicz, L. A., Zafran, J. M., and Brill, R. L. (**2003**). "Discrimination of complex synthetic echoes by an echolocating bottlenose dolphin," J. Acoust. Soc. Am. **113**(2), 1138–1144.

Johnson, S. C. (**1967**). "Sound detection thresholds in marine mammals," in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, New York), Vol. **2**, pp. 247–260.

Plop, R., and Bouman, M. A. (**1959**). "Relation between hearing threshold and duration for tone pulses," J. Acoust. Soc. Am. **31**, 749–758.

Nachtigall, P. E. (**1980**). "Odontocete echolocation performance of object size, shape, and material," in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 71–95.

Schusterman, R. J. (**1980**). "Behavioral methodology in echolocation by marine mammals," in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 11–41.

Urick, R. J. (**1983**). *Principles of Underwater Sound*, 3rd ed. (McGraw–Hill, New York).

Zwislocki, J. (**1960**). "Theory of temporal auditory summation," J. Acoust. Soc. Am. **32**(8), 1046–1060.

Underwater localization of pure tones by harbor seals (*Phoca vitulina*)

Anaïs Bodson, Lars Miersch, and Guido Dehnhardt[a)]
General Zoology & Neurobiology, University of Bochum, ND 6/33, D-44780 Bochum, Germany

The underwater sound localization acuity of harbor seals (*Phoca vitulina*) was measured in the horizontal plane. Minimum audible angles (MAAs) of pure tones were determined as a function of frequency from 0.2 to 16 kHz for two seals. Testing was conducted in a 10-m-diam underwater half circle using a right/left psychophysical procedure. The results indicate that for both harbor seals, MAAs were large at high frequencies (13.5° and 17.4° at 16 kHz), transitional at intermediate frequencies (9.6° and 10.1° at 4 kHz), and particularly small at low frequencies (3.2° and 3.1° at 0.2 kHz). Harbor seals seem to be able to utilize both binaural cues, interaural time differences (ITDs) and interaural intensity differences (IIDs), but a significant decrease in the sound localization acuity with increasing frequency suggests that IID cues may not be as robust as ITD cues under water. These results suggest that the harbor seal can be regarded as a low-frequency specialist. Additionally, to obtain a MAA more representative of the species, the horizontal underwater MAA of six adult harbor seals was measured at 2 kHz under identical conditions. The MAAs of the six animals ranged from 8.8° to 11.7°, resulting in a mean MAA of 10.3°.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2775424]

I. INTRODUCTION

One of the most fundamental features of sound source detection is the ability of an animal to estimate source location as a first step in behaving appropriately in response to the sound (Fay and Popper, 2005). Mammals use two general classes of acoustic cues to localize sound sources: monaural and binaural cues. Monaural spectral cues are processed from one ear only and are the result of the direction-dependent filtering of incoming sound waves accomplished by the pinnae (Wightman and Kistler, 1997). They are thought to be important for determining the elevation of the sound source (Gardner and Gardner, 1973; Hebrank and Wright, 1974a, b) and for determining whether a sound source is in front or behind the listener (Langendijk and Bronkhorst, 2002; Zahorik *et al.*, 2006). In order to enhance filtering information, many animals have evolved large, specially shaped outer ears and many also have the ability to turn the external ear to optimize sound perception. In contrast to monaural localization, binaural localization relies on the comparison of auditory input from the two ears. Binaural cues include interaural time differences (ITDs) and interaural intensity differences (IIDs). Both ITDs and IIDs are used for localizing sounds in the horizontal plane and, together, are the basis of the duplex theory of sound localization (Strutt and Rayleigh, 1907). According to this theory, in humans, frequencies below 1500 Hz are localized by means of ITDs while frequencies above that cut-off are localized primarily by IIDs (Gardner, 1962; Mills, 1958; Stevens and Newman, 1936). Intermediate frequencies at which the auditory system shifts between the utilization of ITDs to IIDs produce poorer localization.

Whereas most terrestrial mammals tested so far use both binaural cues, few species appear to rely on only one or the other of these cues (Heffner *et al.*, 2001a). For example, hedgehog (*Hemiechinus auritus*) and big brown bats (*Eptesicus fuscus*) can use binaural intensity, but not binaural time, whereas horses (*Equus caballus*) and pigs (*Sus scrofa*) can use binaural time, but not binaural intensity (Heffner and Heffner, 1986, 1989; Koay *et al.*, 1998; Masterton *et al.*, 1975). Animals that lack the ability to use ITD are unable to localize low frequencies whereas those that cannot use IID are unable to localize high frequencies in the horizontal plane. Recent results on aerial sound localization in a sea lion (*Zalophus californianus*) and a harbor seal (*Phoca vitulina*) are consistent with the duplex theory of sound localization, where best performance occurred at low and high frequencies while poorest performance occurred at intermediate frequencies. In contrast, the elephant seal's (*Mirounga angustirostris*) poorest performance occurred at higher frequencies suggesting an inferior ability to utilize IID (Holt *et al.*, 2005).

Cues such as ITDs and IIDs are highly functional for localizing sounds in air because the physical structure of the head strongly constrains how sounds reach the ears. The situation differs greatly for aquatic mammals. Underwater sound propagation can profoundly affect the nature of the acoustic stimulus, limiting how useful certain cues will be for deriving spatial information (Branstetter and Mercado, 2006). First, sound velocity is 4–4.5 greater under water (depending on the salinity and ambient temperature) than in air. Second, the similarity between the impedance of the skull and the surrounding water diminishes the head's acoustic shadow effect (Hollien, 1973; Shupak *et al.*, 2005). Consequently, during submersion the ITDs and IIDs required for precise sound localization are largely reduced. In this respect, Hollien *et al.*

[a)]Electronic mail: dehnhardt@neurobiologie.rub.de

(1986) proposed that in the underwater environment, the crossover frequency separating the two types of interaural differences for a human head would occur at approximately 6000 Hz.

The most common sound localization measure is the minimum audible angle (MAA), which is the smallest detectable difference between the azimuths of two identical sound sources (Mills, 1958), with smaller MAAs representing better localization ability. Localization ability improves as the stimulus frequency bandwidth increases and more binaural and spectral cues exist in the stimulus, resulting in better sound localization ability for broadband stimuli, such as clicks and noise, compared with tonal stimuli (Middlebrooks and Green, 1991). Aerial sound localization thresholds have been obtained for over 30 different species of mammals, from mice to elephants (Heffner and Heffner, 1998; Heffner et al., 2001b). In air, thresholds for broadband noise range from about 1° for elephants (*Elephas maximus*; Heffner and Heffner, 1982) to more than 20° for horses (*Equus caballus*; Heffner and Heffner, 1984) and cattle (*Bos taurus*; Heffner and Heffner, 1992), and over 30° for house mice (*Mus musculus*; Heffner and Heffner, 2003). Unfortunately, such a huge database does not exist for aquatic species. A bottlenose dolphin distinguished two sound sources as close as 2°–4° on either side of the midline with tones ranging from 6 to 100 kHz. The bottlenose dolphin had a MAA of 0.9° for simulated echolocation click trains centered at 64 kHz (Renaud and Popper, 1975). With human divers, MAAs were found to be 11.3° at 3.5 kHz, 11.5° at 6.5 kHz, and 7.3° with white noise (Feinstein, 1973). Underwater directional hearing in pinnipeds was investigated experimentally on several occasions. Moore and Au (1975) investigated the sound localization acuity of a California sea lion as a function of frequency. The MAAs for the lower frequencies (4° at 1 kHz) were smaller than for the higher frequencies (18° at 16 kHz), but their animal experienced extreme difficulties localizing pure tones at intermediate frequencies; the MAA at 2 kHz was not obtained and the MAA at 4 kHz was 42.4°. In contrast, Gentry (1967) testing intermediate frequencies with a California sea lion as well, could obtain MAAs of 15° and 10° at 3.5 and 6 kHz, respectively. Møhl (1964) studied underwater sound localization in the harbor seal and reported a MAA of 3.1° for a 2 kHz pure tone signal. However, using a click train, Terhune (1974) obtained a MAA for the same species of 9° (±4°). Recently, another harbor seal tested in our lab (Bodson et al., 2006) provided a MAA of 9.8° at 6 kHz. Although these studies constitute a considerable contribution on underwater directional hearing, they do not provide sufficient data to draw a consistent representation of the sound localization acuity of pinnipeds underwater.

The present study extends, clarifies, and completes the previous work on underwater sound localization in pinnipeds, particularly in harbor seals. First, in order to examine the effect of frequency and consequently the role played by ITDs and IIDs under water, the MAAs of two harbor seals were measured in the horizontal plane at frequencies ranging from 0.2 to 16 kHz. Additionally, to raise the degree of generability regarding the underwater sound localization acuity of the species at one frequency, the horizontal underwater MAA of six adult male harbor seals was measured at 2 kHz under identical conditions.

II. MATERIALS AND METHODS

A. Subjects

Six adult male harbor seals resident at the Marine Science Center at Cologne Zoo (Germany) participated in this study. Bill (8 years old) had experience performing sound localization tasks (Bodson et al., 2006). The other experimental animals, Nick (7 years old), Malte (7 years old), Henry (9 years old), and Sam (12 years old) previously served as subjects in studies on olfaction, vision, and hydrodynamic reception. Marco (24 years old), an almost blind seal, had never been used before in behavioral experiments. The seals were housed in a spacious enclosure consisting of two interconnected pools with a total volume of about 1000 m^3 of fresh water and adjacent land parts. In the current study, the seals received approximately 90% of their daily diet (3–4 kg of mixed herring and sprat) during experimental sessions. Tests were carried out once or twice per day, 5 days per week.

B. Experimental setup

Testing was conducted in the 10-m-diam horizontal half circle arena described in Bodson et al. (2006). However, only a restricted area of the half circle (0–50°) was used for data collection. The half circle consisted mainly of a black polyethylene board fixed 80 cm below the water surface. Small holes were drilled every 5° of the board's horizontal midline. A continuous ribbon of waterproof adhesive tape occluded all holes along the entire inner side of the board.

At a depth of 80 cm, a hoop station exactly fitting the animal's muzzle (Fig. 1) was placed at the center point of the half circle, 5 m away from the sound sources. The seals were trained to station properly in the hoop station, so that head movements were restricted during stimulus presentation. Additionally, to control for body position, two vertical stainless steal posts (1 m high, 70 cm apart from each other) were fixed to the ground 0.5 m behind the hoop station. Two response targets were mounted on both sides of the lower jaw station.

Six underwater loudspeakers (DNH's Aqua-30) were mounted at the outer side of the half circle, each behind one of the holes of the board. In order to exclude any visual cues, the transducers and their respective electric cables were covered at the outer side of the board. The positions of transducers were manually changed between experimental sessions.

Transducers were connected via cables to a selector switch and then to a computer. The computer-generated signal was routed to the selector switch that allowed one of the six connected speakers to be manually selected by the experimenter. All electronic instruments (computer, monitor, keyboard, and speaker selector) were housed in a cupboard placed at the edge of the pool. During tests, the experimenter was concealed by the open doors of the cupboard so that unintentional cuing was excluded.
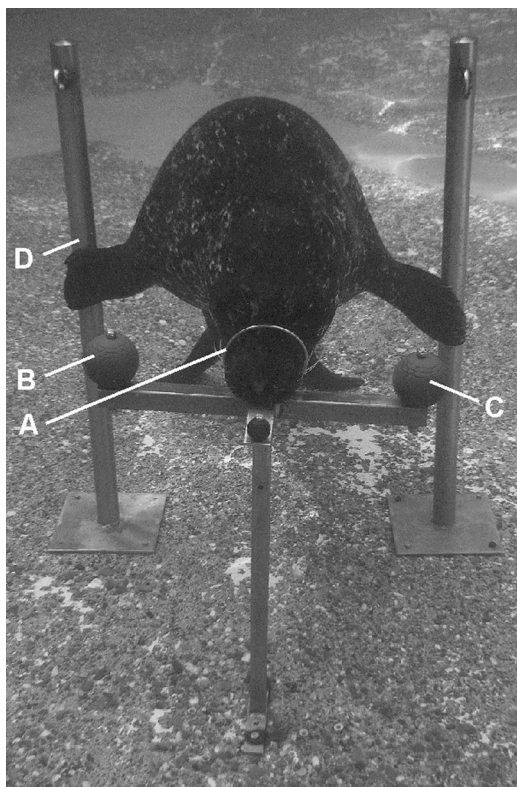
FIG. 1. Harbor seal (Bill) at the station in front of the underwater speakers. (a) Hoop station. (b) Right target response. (c) Left target response. (d) Steal posts.

C. Stimuli

To evaluate the effect of frequency on underwater sound localization of two harbor seals, MAAs were determined for frequencies of 0.2, 0.5, 1, 2, 4, 6, 8, and 16 kHz. The tone duration was 500 ms (10 ms rise and fall times). MAAs of six harbor seals were also determined with a 2 kHz pure tone lasting 500 ms (10 ms rise and fall times).

Since localization ability decreases with decreasing stimulus intensity but remains constant across a range of intensities well above detection threshold (Su and Recanzone, 2001), each frequency was presented at 30 dB above threshold level. The different intensities generated were principally calculated in respect to one underwater audiogram of the harbor seal which covers frequencies between 1 and 180 kHz (Møhl, 1968). The intensity generated at 0.2 kHz was calculated from a low frequency audiogram of the harbor seal (Kastak and Schusterman, 1998). Since the underwater sensitivity of the harbor seal was not available at 0.5 kHz, the threshold at 0.4 kHz (Kastak and Schusterman, 1998) was used instead.

Using a calibrated hydrophone (Brüel & Kjaer, type 8104), the average sound pressure level of stimuli presented from different test angles was determined at the center of the half circle. Differences of up to 3 dB in sound pressure level were observed between stimuli projected from different transducer locations. As data on intensity discrimination in pinnipeds (California sea lion: 3 dB at 16 kHz, Moore and Schusterman, 1976) show that these differences in sound pressure level might be detectable for the seals, the position

of all transducers was randomly switched before every fourth session, so that any differences in sound intensity between transducers was not systematically correlated with a position in the horizontal plane. Since the animal's performances fell to chance level at small angles, it is highly probable that seals could not use intensity differences between speakers or speaker positions to improve performance.

D. Procedure

In order to work in quiet conditions, the seals not involved in an experimental session were separated in another pool and water conduits were switched off before starting the experimental session. For the same reason, no session was conducted while it was raining.

A left/right forced choice procedure was used to determine the underwater MAAs. The seal entered the test area, swam between the posts to the hoop station, and placed its muzzle firmly at the lower jaw station. When the animal was stationed this way, the signal was given. The seal was not allowed to leave the station before the end of the stimulus. The seal was required to respond to a sound from its subjective right or left by pushing the right or left response target, respectively. Each correct response was rewarded by a piece of cut herring. Only one frequency was presented within an experimental session. Prior to testing, the seals were required to demonstrate their ability to perform the task by responding correctly on at least 90% of the trials at angles of 25° for five consecutive sessions. The fastest subject (Bill) achieved this performance criterion by the eighth training session and the slowest (Malte) by the 29th training session (40 trials per session).

The six underwater speakers were positioned at 5°, 10°, 15°, or 20° on either side of a point directly ahead of the animal's station. Within each session, three speakers were placed to the right and three speakers were placed to the left of the subject's midline with a probability of left and right trial presentations of 0.5. The order of presentation was predetermined and pseudorandom (limit of three consecutive sounds from the same position). To obtain one MAA, 50 trials at each test azimuth were collected in 10 sessions (400 trials per MAA).

E. Analysis

To determine the MAAs of our six harbor seals, the percentage of correct choices was converted into percentage of responses to the right target for each of the eight tested angles (Holt et al., 2004, 2005; Mills, 1958). This means that for the four test angles to the right side of the seal's midline, the percentage of right target responses corresponds to the percentage of correct choices. For the four test angles to the left side of the seal's midline (negative angles on the x axis), the percentage of right target responses corresponds to 100% minus the percentage of correct choices. Both linear and sigmoidal regressions were calculated for each psychometric function obtained for different frequencies. The function with the best fit (least error) was chosen to interpolate the respective MAA. The threshold was defined at 75% correct choices, corresponding to 75% right target responses for

TABLE I. Minimum audible angles (MAAs), testing order, and number of sessions needed to reach performance criterion for each harbor seal at each test frequency.

Subject	Test frequency (kHz)	MAA (deg)	Testing order	Number of sessions to reach criterion
Bill	0.2	3.2	8	⋯
	0.5	2.5	7	⋯
	1	4.1	6	⋯
	2	9.4	2	⋯
	4	9.6	3	⋯
	6	9.8	1	8
	8	13	4	⋯
	16	13.5	5	⋯
Marco	0.2	3.1	8	⋯
	0.5	5.0	3	⋯
	1	5.3	2	⋯
	2	11.2	1	13
	4	10.1	6	⋯
	6	8.8	7	⋯
	8	13.4	5	⋯
	16	17.4	4	⋯
Sam	2	8.8	1	20
Malte	2	10.2	1	29
Henry	2	10.4	1	12
Nick	2	11.7	1	10

FIG. 3. Individual variation of underwater sound localization for six harbor seals at 2 kHz. Dots indicate the mean percentage of responses to the right target of the six subjects. Vertical lines illustrate the dispersion of the distribution, with the upper and lower limits indicating the maximum and minimum percentage of responses to the right target.

sound presentations from the right side of the seal's midline and 25% right target responses for sound presentations from the left side of the seal's midline. The MAA corresponds to the mean of the two threshold values.

III. RESULTS

MAAs, frequencies tested, subjects, testing orders, and number of sessions needed to achieve the performance criterion are given in Table I. MAAs as a function of frequency for two harbor seals (Bill and Marco) are illustrated in Fig. 2. The exact MAAs were obtained by means of linear interpolation except for 0.2, 0.5, and 1 kHz where data points followed a sigmoidal curve. In these cases, a sigmoid function was used to interpolate the thresholds. For both subjects, the

FIG. 2. Minimum audible angles (MAAs) of two harbor seals as a function of the frequencies tested.

general run of the curve shows that their sound localization acuity decreased with increasing frequency. For both seals, the smallest MAAs were determined at low frequencies (2.5° at 0.5 kHz for Bill and 3.1° at 0.2 kHz for Marco) and the biggest MAAs at 16 kHz, the highest frequency tested (13.5° and 17.4° for Bill and Marco, respectively). MAAs were transitional at intermediate frequencies (9.4°, 9.6°, 9.8° and 11.2°, 10.1°, 8.8° at 2, 4, 6 kHz for Bill and Marco, respectively). Results could be categorized in three classes: a low-frequency range producing the best underwater sound localization acuity (frequency ≤ 1 kHz, MAA ≤ 5.3°), a frequency range of intermediate sound localization acuity (2–6 kHz, MAAs approximately 10°) and a high-frequency range producing the poorest underwater sound localization acuity (frequency ≥ 8 kHz, MAA ≥ 13°).

MAAs of the six male harbor seals tested at 2 kHz ranged from 8.8° (Sam) to 11.7° (Nick). The six values are equally distributed and their average corresponds to 10.3°. The individual variations in performance are shown in Fig. 3. The sound localization acuity of the experimental subjects was not correlated with parameters such as their age, the number of sessions needed to achieve the performance criterion, their experimental experience, or their visual acuity. As an example, the MAA of the old seal Marco, who was experimentally naive and who is almost blind, is comparable to those of younger, experienced, and good sighted seals.

IV. DISCUSSION

The results of this experiment demonstrate that harbor seals can localize a range of pure tones under water. Their ability to localize low-frequency as well as high-frequency pure tone signals suggests that they are capable of using both time differences and intensity differences. Thus, like most other mammals, harbor seals likely make use of the two primary binaural localization cues, ITDs and IIDs.

The sound localization acuity of the seals varied with the frequency of the stimulus. Specifically, MAAs obtained at

Bodson *et al.*: Underwater sound localization by harbor seals

the three lowest frequencies, 200 Hz, 500 Hz, and 1 kHz, were consistently smaller than those obtained at any of the higher frequencies. Frequencies in the midrange, 2, 4, and 6 kHz, although not localized as well as the lower frequencies, were localized with greater accuracy than higher frequencies. Finally, MAAs obtained at the two highest frequencies, 8 and 16 kHz, were definitely the largest. Studies on the localization of tones suggest that ITDs and IIDs are equally accurate for human subjects (Mills, 1958), while in most other mammals one cue may be less accurate, and perhaps less significant than the other (Brown and May, 2005). As an example, the ability of an Asian elephant (*Elephas maximus*) (Heffner and Heffner, 1982) and an elephant seal (Holt *et al.*, 2005) to localize pure tones in air demonstrates that they are capable of using both cues to localize sound. However, their poor performance in localizing high-frequency tones suggests that both species rely more heavily on the time cue than on the intensity cue for accurate sound localization. This trend seems to be valid for harbor seals under water as well. Indeed, a significant decrease in the sound localization acuity of Bill and Marco with increasing frequency suggests that IID cues may not be as robust as ITD cues.

The MAA of 9° (±4°) of a harbor seal measured under water with a click train by Terhune (1974) is in good accordance with the MAAs obtained in the present study at intermediate frequencies. In contrast, the MAA of the harbor seal tested by Møhl (1964) is about three times smaller (3.1° at 2 kHz) than the mean MAA of 10.3° determined in the present study for six animals at the same frequency. In part, the present results obtained for harbor seals are similar to those obtained by Moore and Au (1975) for the California sea lion. Indeed, for both species, MAAs for lower frequencies were smaller than for higher frequencies, supporting the idea that under water pinnipeds, or at least harbor seals and sea lions, use ITD cues more efficiently than IID cues. In contrast to the California sea lion tested by Moore and Au (1975), which experienced extreme difficulties in localizing sounds at intermediate frequencies (MAA at 2 kHz was not obtained due to the animal's frustration and MAA at 4 kHz was 42.4°), the six harbor seals of this study show normal performance at 2 kHz (MAA ranged from 8.8° to 11.7°). As Gentry (1967) determined a MAA (15° at 3.5 kHz) approximately three times smaller than Moore and Au (1975), the acuity of underwater sound localization of California sea lions at intermediate frequencies is still vague.

A complementary study investigating the localization of aerial pure tones was done with three species of pinnipeds including a harbor seal (Holt *et al.*, 2005). The comparison of their results with those from the present study suggests that harbor seals localize tones better in air than under water except for frequencies below 2 kHz, where performance is similar in both media. That sound localization is easier in air than under water for harbor seals was already shown by Terhune (1974) and can be explained simply by the fact that ITDs and IIDs are largely reduced under water. The novelty here is that despite this fact, the sound localization acuity of harbor seals at low frequencies is at least as good in water than in air. Even though the pinniped ear is not acoustically

isolated from the skull as it is in cetaceans (Repenning, 1972), the underwater MAAs obtained at 200 Hz (3.2° and 3.1° for Bill and Marco, respectively) are similar to those obtained with a bottlenose dolphin for sounds from 6 to 100 kHz (between 2° and 4°, Renaud and Popper, 1975). Thus, the harbor seal seems to be a low frequency specialist for localizing sounds under water, although the sensory mechanism underlying this ability still has to be clarified.

It is known that most behavioral studies of marine mammal sensory systems as well as those of other large species are limited to small sample sizes. Consequently, some questions concerning the confidence or general applicability that can be attributed to these findings remain (Terhune and Turnbull, 1995; Watkins and Wartzok, 1985). Therefore, one objective of this study was to raise the degree of generability regarding the underwater sound localization acuity of the harbor seal. First of all, the sample size was increased to six individuals of the species. Additionally, to obtain the most representative value of the underwater sound localization acuity of harbor seals, every animal was intensively trained to reach a high performance criterion and to keep it constant prior to data collection (minimum 90% correct choices for five consecutive sessions). Thus, all experimental animals started the testing period at the same performance level so that the resulting MAA value should not be affected by the individual stage of learning. The results showed that the six underwater MAAs obtained at 2 kHz are indeed very similar. It is clear that the examination of six well-trained harbor seals does not characterize fully the sound localization acuity of the species, but at least it provides a more confident value of their MAA at 2 kHz.

The sound localization acuity of the severely visually impaired seal Marco was neither better nor worse than the five normal sighted seals at 2 kHz. Additionally, considered as a function of frequency, his MAAs were similar to those of Bill, some slightly better and some slightly worse. According to Brown and May (2005), there subsist two contrasting perspectives concerning the effect of blindness on sound localization. On one hand, it is possible that spatial maps are organized by visual experience. Thus visual deprivation would lead to impairments in sound localization. On the other hand, it is suggested that visual deprivation might produce compensatory sharpening of directional hearing. It was confirmed that blind persons show a clear superiority at higher levels of the auditory pathways as compared to a control group (Niemeyer and Starlinger, 1981). Results from physiological studies are consistent with the idea that early blindness may result in cross-modal reorganization of the cortex, and this reorganization may produce compensatory effects for sound localization (Kujala *et al.*, 1992, 1995, 2000). Behavioral data consistent with this hypothesis have been reported for ferrets (King and Parsons, 1999) and cats (Rauschecker and Harris, 1983; Rauschecker and Kniepert, 1994). Since Marco showed normal MAAs irrespective of the frequency tested, his case does not favor one or the other of these perspectives. His normal acuity of sound localiza-

tion could be explained by the fact that he is not totally blind and was probably not deprived of vision at early stages of his development.

The ability of harbor seals to localize a range of different frequencies might be useful in many underwater activities such as hunting soniferous fish, avoiding predators, or simply orientating and navigating in a noisy ocean. In a social context, good sound localization abilities might play an important role between mothers and pups during the lactation period. A harbor seal pup vocalizes almost continuously when following its mother, emitting a call that is transmitted in air and underwater simultaneously. These calls presumably facilitate the contact between mother and pup as well as their reunion once separated (Renouf, 1984). Furthermore, the harbor seal is an aquatic-mating phocid with males producing underwater vocalizations consisting of low-frequency broadband growls or "roars" (Hanggi and Schusterman, 1994) with most energy concentrated around 280 Hz (±74 Hz) (Bjørgesaeter et al., 2004). These vocalizations occur mainly during the mating season (Van Parijs et al., 1997) and may be used for advertising male presence to females (Van Parijs et al., 1999). Additionally, they may play a role in intrasexual dominance hierarchies (Hayes et al., 2004). As shown in this study, the underwater sound localization acuity of harbor seals is exceptionally good at 200 Hz (MAA of 3.2° and 3.1° for Bill and Marco, respectively), the tested frequency coming closest to that of their own vocalizations.

Although our experimental pool (about 1000 m³) is bigger than most other tanks used to collect data on marine mammal hearing, we cannot ignore the fact that it likely constitutes a reverberant environment. When a sound is produced in a reverberant environment, it propagates in multiple directions and is subsequently reflected from nearby surfaces. The auditory system is thus faced with resolving competition between the first sound and its reflections for perception and localization (Litovski et al., 1999). However, it has been demonstrated that judgments of localization are determined almost exclusively by the interaural cues carried by the earlier, or direct, sound (Wallach et al., 1949). The observation that it is the first information arriving at the ears which dominates the perception of sound location is referred to as the "precedence effect" and can explain why the ability to localize the source of a sound is remarkably unimpaired by the presence of reflections (Wallach et al., 1949; Yost and Soderquist, 1984; Zurek, 1980). Nevertheless, it would be worth comparing data on marine mammal sound localization acuity collected in experimental tanks with their directional hearing ability in a deep open-water environment.

ACKNOWLEDGMENTS

Bjørgesaeter, A., Ugland, K. I., and Bjørge, A. (**2004**). "Geographic variation and acoustic structure of the underwater vocalization of harbor seal (*Phoca vitulina*) in Norway, Sweden and Scotland," J. Acoust. Soc. Am. **116**, 2459–2468.

Bodson, A., Miersch, L., Mauck, B., and Dehnhardt, G. (**2006**). "Underwater sound localization by a swimming harbor seal," J. Acoust. Soc. Am. **120**, 1550–1557.

Branstetter, B. K., and Mercado, E., III (**2006**). "Sound localization by cetaceans," Int. J. Comp. Psychol. **19**, 26–61.

Brown, C. H., and May, B. J. (**2005**). "Comparative mammalian sound localization," in *Sound Source Localization*, edited by A. N. Popper and R. R. Fay (Springer, New York), pp. 124–178.

Fay, R. R., and Popper, A. N. (**2005**). "Introduction to sound source localization," in *Sound Source Localization*, edited by A. N. Popper and R. R. Fay (Springer, New York), pp. 1–5.

Feinstein, S. H. (**1973**). "Acuity of the human sound localization response underwater," J. Acoust. Soc. Am. **53**, 393–399.

Gardner, M. B. (**1962**). "Binaural detection of single frequency signals in the presence of noise," J. Acoust. Soc. Am. **34**, 1824–1830.

Gardner, M. B., and Gardner, R. S. (**1973**). "Problem of localization in the median plane: Effect of pinnae cavity occlusion," J. Acoust. Soc. Am. **53**, 400–408.

Gentry, R. L. (**1967**). "Underwater auditory localization in the California sea lion (*Zalophus californianus*)," J. Aud Res. **7**, 187–193.

Hanggi, E. B., and Schusterman, R. J. (**1994**). "Underwater acoustic displays and individual variation in male harbour seals, *Phoca vitulina*," Anim. Behav. **48**, 1275–1283.

Hayes, S. A., Kumar, A., Costa, D. P., Mellinger, D. K., Harvey, J. T., Southall, B. L., and Le Boeuf, B. J. (**2004**). "Evaluating the function of the male harbour seal, *Phoca vitulina*, roar through playback experiments," Anim. Behav. **67**, 1133–1139.

Hebrank, J., and Wright, D. (**1974a**). "Are two ears necessary for localization of sound sources on the median plane?" J. Acoust. Soc. Am. **56**, 935–938.

Hebrank, J., and Wright, D. (**1974b**). "Spectral cues used in the localization of sound sources on the median plane," J. Acoust. Soc. Am. **56**, 1829–1834.

Heffner, H. E., and Heffner, R. S. (**1984**). "Sound Localization in large mammals: localization of complex sounds by horses," Behav. Neurosci. **98**, 541–555.

Heffner, H. E., and Heffner, R. S. (**1998**). "Hearing," in *Comparative Psychology, a Handbook*, edited by G. Greenberg and M. M. Haraway (Garland, New York), pp. 290–303.

Heffner, H. E., and Heffner, R. S. (**2003**). "Audition," in *Handbook of Research Methods in Experimental Psychology*, edited by S. Davis (Blackwell Publishing, Malden, Massachusetts), pp. 413–440.

Heffner, R. S., and Heffner, H. E. (**1982**). "Hearing in the elephant (*Elephas maximus*): Absolute sensitivity, frequency discrimination, and sound localization," J. Comp. Physiol. Psychol. **96**, 926–944.

Heffner, R. S., and Heffner, H. E. (**1986**). "Localization of tones by horses: Use of binaural cues and the role of the superior olivary complex," Behav. Neurosci. **100**, 93–103.

Heffner, R. S., and Heffner, H. E. (**1989**). "Sound localization, use of binaural cues and the superior olivary complex in pigs," Brain Behav. Evol. **33**, 248–258.

Heffner, R. S., and Heffner, H. E. (**1992**). "Hearing in large mammals: Sound-localization acuity in cattle (*Bos taurus*) and goats (*Capra hircus*)," J. Comp. Physiol. Psychol. **106**, 107–113.

Heffner, R. S., Koay, G., and Heffner, H. E. (**2001a**). "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.

Heffner, R. S., Koay, G., and Heffner, H. E. (**2001b**). "Sound localization acuity changes with age in C57BL/6J mice," in *Handbook of Mouse Auditory Research: From Behavior to Molecular Biology*, edited by J. F. Willott (CRC Press, Boca Raton, FL), pp. 31–35.

Hollien, H. (**1973**). "Underwater sound localization in humans," J. Acoust. Soc. Am. **53**, 1288–1293.

Hollien, H., Hicks, J. W., and Klepper, B. (**1986**). "An acoustic approach to diver navigation," Undersea Biomed. Res. **13**, 111–128.

Holt, M. M., Schusterman, R. J., Kastak, D., and Southall, B. L. (**2005**). "Localization of aerial pure tones by pinnipeds," J. Acoust. Soc. Am. **118**, 3921–3926.

Holt, M. M., Schusterman, R. J., Southall, B. L., and Kastak, D. (**2004**). "Localization of aerial broadband noise by pinnipeds," J. Acoust. Soc. Am. **115**, 2339–2345.

Kastak, D., and Schusterman, R. J. (**1998**). "Low-frequency amphibious hearing in pinnipeds: Methods, measurements, noise, and ecology," J.

Acoust. Soc. Am. **103**, 2216–2228.

King, A. J., and Parsons, C. H. (**1999**). "Improved auditory spatial acuity in visually deprived ferrets," Eur. J. Neurosci. **11**, 3945–3956.

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.

Kujala, T., Alho, K., Kekoni, J., Hamalainen, H., Reinikainen, K., Salonen, O., Standertskjold-Nordenstam, C. G., and Näätänen, R. (**1995**). "Auditory and somatosensory event-related brain potentials in early blind humans," Exp. Brain Res. **104**, 519–526.

Kujala, T., Alho, K., and Näätänen, R. (**2000**). "Cross-modal reorganization of human cortical functions," Trends Neurosci. **23**, 115–120.

Kujala, T., Alho, K., Paavilainen, P., Summala, H., and Näätänen, R. (**1992**). "Neural plasticity in processing of sound location by the early blind: An event-related potential study," Electroencephalogr. Clin. Neurophysiol. **84**, 469–472.

Lagendijk, E. H., and Bronkhorst, A. W. (**2002**). "Contribution of spectral cues to human sound localization," J. Acoust. Soc. Am. **112**, 1583–1596.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.

Masterton, B., Thompson, G. C., Bechtold, J. K., and Robards, M. J. (**1975**). "Neuroanatomical basis of binaural phase difference analysis for sound localization: A comparative study," J. Comp. Physiol. Psychol. **89**, 379–386.

Middlebrooks, J. C., and Green, D. M. (**1991**). "Sound localization by human listeners," Annu. Rev. Psychol. **42**, 135–159.

Mills, A. W. (**1958**). "On the minimum audible angle," J. Acoust. Soc. Am. **30**, 237–246.

Møhl, B. (**1964**). "Preliminary studies on hearing in seals," Vidensk. Medd. Dansk. Naturh. Foren. **127**, 283–294.

Møhl, B. (**1968**). "Auditory sensitivity of the common seal in air and water," J. Aud Res. **8**, 27–38.

Moore, P. W. B., and Au, W. L. (**1975**). "Underwater localization of pulsed pure tones by the California sea lion (*Zalophus californianus*)," J. Acoust. Soc. Am. **58**, 721–727.

Moore, P. W. B., and Schusterman, R. J. (**1976**). "Discrimination of pure-tone intensities by the California sea lion," J. Acoust. Soc. Am. **60**, 1405–1407.

Niemeyer, W., and Starlinger, I. (**1981**). "Do the blind hear better? Investigations on auditory processing in congenital or early acquired blindness. II. Central functions," Audiology **20**, 510–515.

Rauschecker, J. P., and Harris, L. R. (**1983**). "Auditory compensation of the effects of visual deprivation in the cat's superior colliculus," Exp. Brain Res. **50**, 69–83.

Rauschecker, J. P., and Kniepert, U. (**1994**). "Auditory localization behav-iour in visually deprived cats," Eur. J. Neurosci. **6**, 149–160.

Renaud, D. L., and Popper, A. N. (**1975**). "Sound localization by the bottle-nose porpoise *Tursiops truncatus*," J. Exp. Biol. **63**, 569–585.

Renouf, D. (**1984**). "The vocalization of the harbour seal pup (*Phoca vitulina*) and its role in the maintenance of contact with the mother," J. Zool. **202**, 583–590.

Reppening, C. A. (**1972**). "Underwater hearing in seals: Functional morphology," in *Functional Anatomy of Marine Mammals*, edited by R. J. Harrison (Academic, London), Vol. **I.**, pp. 307–331.

Shupak, A., Sharoni, Z., Yanir, Y., Keynan, Y., Alfie, Y., and Halpern, P. (**2005**). "Underwater hearing and sound localization with and without an air interface," Otol. Neurotol. **26**, 127–130.

Stevens, S. S., and Newman, E. B. (**1936**). "The localization of actual sources of sound," Am. J. Psychol. **48**, 297–306.

Strutt, J. W., and Rayleigh, L. (**1907**). "On our perception of sound direction," Philos. Mag. **13**, 214–232.

Su, T. I. K., and Recanzone, G. H. (**2001**). "Differential effect of near-threshold stimulus intensities on sound localization performance in azimuth and elevation in normal human subjects," J. Assoc. Res. Otolaryngol. **2**, 246–256.

Terhune, J. M. (**1974**). "Directional hearing of a harbor seal in air and water," J. Acoust. Soc. Am. **56**, 1862–1865.

Terhune, J. M., and Turnbull, S. (**1995**). "Variation in the psychometric functions and hearing thresholds of a harbour seal," in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden, The Netherlands), pp. 81–93.

Van Parijs, S. M., Hastie, G. D., and Thompson, P. M. (**1999**). "Geographic variation in temporal and spatial vocalization patterns of male harbour seals in the mating season," Anim. Behav. **58**, 1231–1239.

Van Parijs, S. M., Thompson, P. M., Tollit, D. J., and Mackay, A. (**1997**). "Distribution and activity of male harbour seals during the mating season," Anim. Behav. **54**, 35–43.

Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**). "The precedence effect in sound localization," Am. J. Psychol. **62**, 315–336.

Watkins, W. A., and Wartzok, D. (**1985**). "Sensory biophysics of marine mammal," Marine Mammal Sci. **1**, 219–260.

Wightman, F. L., and Kistler, D. J. (**1997**). "Monaural sound localization revisited," J. Acoust. Soc. Am. **101**, 1050–1063.

Yost, W. A., and Soderquist, D. R. (**1984**). "The precedence effect: Revisited," J. Acoust. Soc. Am. **76**, 1377–1383.

Zahorik, P., Bangayan, P., Sundareswaran, V., Wang, K., and Tam, C. (**2006**). "Perceptual recalibration in human sound localization: Learning to remediate front-back reversals," J. Acoust. Soc. Am. **120**, 343–359.

Zurek, P. M. (**1980**). "The precedence effect and its possible role in the avoidance of interaural ambiguities," J. Acoust. Soc. Am. **67**, 952–964.

Asymmetric airflow and vibration induced by the Coanda effect in a symmetric model of the vocal folds

Chao Tao, Yu Zhang, Daniel G. Hottinger, and Jack J. Jiang[a]
Department of Surgery, Division of Otolaryngology Head and Neck Surgery, University of Wisconsin
Medical School, Madison, Wisconsin 53792-7375

A model constructed from Navier-Stokes equations and a two-mass vocal fold description is proposed in this study. The composite model not only has the capability to describe the aerodynamics in a vibratory glottis but also can be used to study the vocal fold vibration under the driving of the complex airflow in the glottis. Numerical simulations show that this model can predict self-oscillations of the coupled glottal aerodynamics and vocal fold system. The Coanda effect could occur in the vibratory glottis even though the vocal folds have left-right symmetric prephonatory shape and tissue properties. The Coanda effect causes the asymmetric flow in the glottis and the difference in the driving force on the left and right vocal folds. The different pressures applied to the left and right vocal folds induce their displacement asymmetry. By using various lung pressures (0.6–2.0 kPa) to drive the composite model, it was found that the asymmetry of the vocal fold displacement is increased from 1.87% to 11.2%. These simulation results provide numerical evidence for the presence of asymmetric flow in the vibratory glottis; moreover, they indicate that glottal aerodynamics is an important factor in inducing the asymmetric vibration of the vocal folds. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2773960]

I. INTRODUCTION

During phonation, it is continual energy transfer, through the interaction of airflow and tissue, that drives vocal fold oscillation and phonation (Titze, 1994). Vocal fold properties (i.e., vocal fold mass, stiffness, dehydration, etc.) as well as glottal aerodynamics affect phonation. Therefore, in order to understand voice production, it is important to study not only laryngeal physiology and tissue characteristics but also the aerodynamics of the laryngeal airway.

It has been found that the airflow stream in the glottis has the tendency to stay attached to one side of the vocal fold, rather than follow a straight line in its original direction. This is the so-called Coanda effect in aerodynamics. Some investigators have suggested that Coanda effects in the glottis could play an important role in vocal fold vibration and voice production. Teager and Teager (1983), Kaiser (1983), and Pelorson *et al.* (1994) pointed out that Coanda effects could induce supraglottal airflow asymmetry. Pelorson and Hirschberg (1997) also presented simultaneous pressure measurements to confirm the occurrence of a Coanda effect. Hofmans *et al.* (2003) showed the asymmetry of the flow in the glottis due to the Coanda effect by using *in vitro* models. Shinwari *et al.* (2003) found that even for a symmetric glottis, the glottal airflow distribution could be asymmetric. Recently, Erath and Plesniak (2006a, b, c) also studied the Coanda effect in pulsatile flow through a scaled-up vocal fold model. It was found the attachment of the flow to the vocal-fold walls occurred for divergence angles of less than 20°. The asymmetric pressure distribution on left and right

vocal folds arising from asymmetric flow could also result in a difference in the aerodynamic forces acting on the two folds. Therefore, it was inferred that the presence of Coanda effects in the glottis could influence vocal fold vibration due to the drastic change in the separation point of the flow.

The influences of Coanda effects on vocal fold vibration have not been well examined because in past studies the interaction between the airflow and vocal fold is often ignored. It is known that voice production is a complicated process including aerodynamics and biomechanics. The pressure applied to the vocal fold surface is the driving force for vocal fold vibration. The interaction between airflow and vocal fold generates the vibration of the vocal fold. However, this airflow-tissue interaction cannot be effectively described by oft-used static models, where a fixed airflow boundary is used instead of the actual vocal fold movement (Iijima *et al.*, 1992; Guo and Scherer, 1993; Alipour *et al.*, 1996; Hofmans *et al.*, 2003; Scherer *et al.*, 2001, 2002; Alipour and Scherer, 2002; Shinwari *et al.*, 2003; Kucinschi *et al.*, 2006; Erath and Plesniak, 2006a, b, c). The forced-oscillating model moves the vocal folds by sinusoidal medial-lateral oscillation to simulate the changing glottal area. Although the force-oscillating model may represent a better approximation than the static model (Zhang *et al.*, 2002; Zhao *et al.*, 2002; Alipour and Schere, 2004; Erath and Plesniak, 2006a, b, c), the fluid-tissue interaction in these models is unidirectional and the effect of the airflow on vocal fold movement is still not considered. Therefore, these models could be inadequate for studying the influence of the Coanda effect on vocal fold vibration during phonation.

In addition, the mere presence of the Coanda effect in a vibratory glottis is still arguable, because most of the aforementioned studies are based on static physical and numerical

a)Electronic mail: jjjiang@wisc.edu

models. Hofmans *et al.* (2003) found that Coanda effects take a long time to develop in these static glottal models and they suggested that the Coanda effect is unlikely to occur during normal speech production. In contrast, Shinwari *et al.* (2003) pointed out that the glottis begins to take on a divergent shape only near the peak of the glottal flow, when the Coanda effect may have the best chance to form. Moreover, Hofmans (1998) showed that once the volume flow acceleration decreases sufficiently, the Coanda effect takes place very quickly (within 1 to 2 ms). These studies suggested that the Coanda effect could occur during phonation (Shinwari *et al.*, 2003). It would be difficult, therefore, for the static models to provide direct evidence for the presence of the Coanda effect and asymmetric flow in a vibratory glottis during phonation.

In order to be able to provide a primary answer to these pending questions, a new self-oscillating glottal model that includes airflow-tissue interaction is proposed in this study. In this model, the glottal aerodynamics are described by the Navier-Stokes (NS) fluid equations and solved with the finite-element technique, while the boundary condition is solved from the interaction between airflow and a two-mass vocal fold using the fourth-order Runge-Kutta routine. Because the aerodynamics part and the vocal fold part of this model are described by equations with different types (partial differential equations and differential equations) and solved with different numerical techniques, we call the model a composite model. In comparison to those models based on the highly idealized Bernoulli's law (Ishizaka and Flanagan, 1972; Titze, 1988; Lucero, 1996; Jiang *et al.*, 2001; Tao and Jiang, 2006), the composite model can provide a more vivid description of the aerodynamics in the vibratory glottis. In comparison to full continuous models, where both aerodynamics and vocal fold are modeled with the finite-element or finite-difference scheme (Alipour *et al.*, 1996; Alipour and Titze, 1996; De Oliveira Rosa *et al.*, 2003; Thomson *et al.*, 2005; Tao *et al.*, 2006), the lumped mass parameters of this composite model may not directly represent the physiological properties of vocal fold tissues, however, this model provides a good physical veracity of glottal aerodynamics with less computational cost. In comparison to a static boundary model and force-oscillating model, this model not only can predict the aerodynamics in a vibratory glottis during phonation, but also can predict the vocal fold vibration induced by the complex aerodynamics. The detailed development of this composite model will be presented in Sec. II. In Sec. III we will compare this model with a Bernoulli-based two-mass model; we then use this composite model to investigate the asymmetric velocity distribution and asymmetric pressure distribution in the vibratory glottis and study the asymmetric vocal fold vibration induced by the asymmetric glottal airflow. Finally, we draw conclusions from this analysis.

II. METHOD

In this study, the airflow was assumed to be incompressible Newtonian fluid. The acoustic wave propagation was ignored in this airflow model. The behavior of the flow field was defined by the NS equations and the movement of the glottal wall was solved from a two-mass model. The details of this model development are also given.

A. Design of the airflow channel

A two-dimensional computational fluid dynamics (CFD) code was used to predict the aerodynamics in the glottis. Usually, three laws of conservation (mass, momentum, and energy) are needed to describe an airflow problem. However, in this study, based upon the assumption of an incompressible Newtonian fluid in the glottis, the energy equation is decoupled from the flow. Therefore, an incompressible flow can be described by the Navier-Stokes equations (which govern momentum) and the continuity equation (which governs mass conservation). The law of conservation of mass that can be governed by the continuity equation is:

$$\nabla \cdot \mathbf{V} = 0, \tag{1a}$$

where ρ is flow density and $\mathbf{V}$ is the velocity vector of flow. The momentum equation is

$$\rho \frac{\partial \mathbf{V}}{\partial t} + \rho \mathbf{V} \cdot \nabla \mathbf{V} = \rho \mathbf{g} - \nabla p + \mu \nabla^2 \mathbf{V}, \tag{1b}$$

where p is pressure and μ is viscosity. $\mathbf{g}$ represents acceleration due to gravity, which was ignored in this study. In the following simulation, the density of the fluid is 1.2 kg/m^3 and the viscosity of the fluid is $1.8 \times 10^{-5} \text{ N s/m}^2$, which are close to the properties of air at $37\,^\circ$C. The airflow channel was defined as follows: The glottal length was 0.3 cm, and both the prephonatory superior and inferior glottal widths were 0.17857 mm. A 2 cm inlet duct and an 8 cm outlet duct represented the trachea and vocal tract, respectively. Their widths were both 2 cm. The outline of glottal inlet can be roughly fitted by the polynomial $\pm x = -0.018 + 0.519y - 0.828y^2 + 4.351y^3 + 8.524y^4 + 3.809y^5$ $(-1.0 < y < 0.0)$ and the outline of glottal outlet can be roughly fitted by the polynomial $\pm x = -11 + 117y - 439y^2 + 765y^3 - 638y^4 + 207y^5 (0.3 < y < 0.8)$ for the glottal outlet. Figure 1 presents the structure of the airflow channel around the glottis.

The above-mentioned partial differential equation [Eq. (1)] has to be completed by boundary conditions in the glottis. The no-slip condition was applied on the tracheal wall and the vocal tract wall, that is, all the velocity components were set to zero. The specified pressure condition was applied on the inlet of the subglottal tube to drive the model, meaning that the pressure at the inlet was specified as the constant lung pressure. The zero pressure boundary condition was set at the outlet of the supraglottal tube. A moving wall condition was applied on the fluid-solid interaction boundary, i.e., the displacement and velocity at the interaction surface is specified as the values according to the solution of a two-mass model. The details of solving a two-mass model and applying a moving wall boundary condition will be introduced in Sec. II B. With the given boundary condition, the NS equations were solved completely.
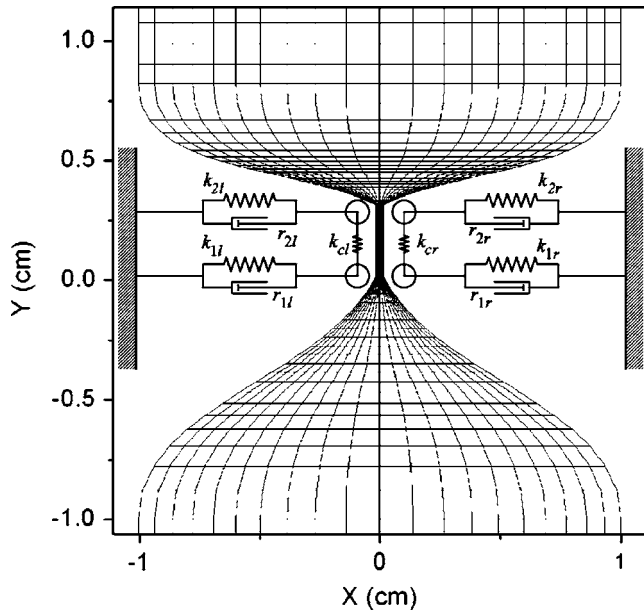
FIG. 1. The sketch map of the composite glottal model. The glottal aerodynamics are described by the Navier-Stokes equation and solved with the finite-element technique, while the vocal fold is described by a two-mass model with fourth-order Runge-Kutta routine.

B. The glottal wall movement

The two-mass model has been used to study many interesting phenomena such as the mucosal wave propagation, chaos, and other irregular vibration patterns. Moreover, compared with the finite-element model, the two-mass model is a simple and effective way to describe vocal fold vibration. The two-mass model can be described as follows (Steinecke and Herzel, 1995):

$$\dot{x}_{1\alpha} = v_{1\alpha},$$

$$\dot{v}_{1\alpha} = \frac{1}{m_{1\alpha}}\left(P_{1\alpha}Ld_{1\alpha} - r_{1\alpha}v_{1\alpha} - k_{1\alpha}x_{1\alpha} - \Theta(-a_{1\alpha})c_{1\alpha}\frac{a_{1\alpha}}{2L}\right.$$
$$\left. - k_{c\alpha}(x_{1\alpha} - x_{2\alpha})\right), \tag{2}$$

$$\dot{x}_{2\alpha} = v_{2\alpha},$$

$$\dot{v}_{2\alpha} = \frac{1}{m_{2\alpha}}\left(-r_{2\alpha}v_{2\alpha} - k_{2\alpha}x_{2\alpha} - \Theta(-a_{2\alpha})c_{2\alpha}\frac{a_{2\alpha}}{2L}\right.$$
$$\left. - k_{c\alpha}(x_{2\alpha} - x_{1\alpha})\right), \tag{3}$$

where $x_{i\alpha}$ is the displacement of mass $m_{i\alpha}$, $r_{i\alpha}$ is the damping constant, $k_{i\alpha}$ is the spring constant, $d_{1\alpha}$ is the thickness of mass $m_{1\alpha}$, $L = 1.4$ cm is the length of the glottis, $c_{i\alpha}$ is the additional spring constant during collision, and $k_{c\alpha}$ is the coupling constant between the upper mass and lower mass. The index i denotes the inferior ($i=1$) and superior masses. The index α represents right ($\alpha=r$) and left ($\alpha=l$) masses. $a_1 = a_{01} + (x_{1r} + x_{1l})L$ is the lower glottal area and $a_2 = a_{02} + (x_{2r} + x_{2l})L$ is the upper glottal area, where a_{0i} is the prephonatory lower ($i=1$) and upper ($i=2$) glottal areas.

TABLE I. Parameter values for the lumped-mass model.

Parameter		Value
$m_{1\alpha}$	Lower mass	0.125 g
$m_{2\alpha}$	Upper mass	0.025 g
$k_{1\alpha}$	Lower spring constant	80 kdyn/cm
$k_{2\alpha}$	Upper spring constant	8 kdyn/cm
$k_{c\alpha}$	Coupling spring constant	25 kdyn/cm
$r_{1\alpha}$	Damping constant	20 g/s
$r_{2\alpha}$	Damping constant	20 g/s

The function $\Theta(x)$ describing the collision is approximated by

$$\Theta(x) = \begin{cases} \tanh(50x/x_0), & x > 0 \\ 0, & x \le 0. \end{cases} \tag{4}$$

When the glottis is open ($a_1 \ge 0$ and $a_2 \ge 0$), the function $\Theta(x)$ with $x \le 0$ is zero. The term $\Theta(-a_{i\alpha})c_{i\alpha}a_{i\alpha}/2L$ used to represent collision is ignored in Eqs. (2) and (3). When the glottis is closed and collision occurs ($a_1 < 0$ or $a_2 < 0$), the function $\Theta(x)$ with $x > 0$ is positive, moreover, $\Theta(x) \to 1$ if $x \gg 1$. The collision term $\Theta(-a_{i\alpha})c_{i\alpha}a_{i\alpha}/2L$ in Eqs. (2) and (3) is activated to describe the vocal fold collision.

$P_{1\alpha}$ is the average airflow pressure applied on the mass $m_{1\alpha}$. When the glottis is open ($a_1 \ge 0$ and $a_2 \ge 0$), $P_{1\alpha}$ can be integrated from the pressure solution of the NS equations (Titze, 1988):

$$P_{1\alpha} = \frac{1}{D}\int_0^D p_\alpha(y)dy, \tag{5}$$

where D is the thickness of vocal fold and $p_\alpha(y)$ is the pressure on the vocal glottis wall solved from NS equations. When collision occurs ($a_1 < 0$ or $a_2 < 0$), the airflow is interrupted by the closed glottis. The air pressure below the collision point is equal to the lung pressure and the air pressure above the collision point is equal to the reference pressure at the outlet of the supraglottal tube. In Eqs. (2) and (3), these lumped-mass parameters, such as $m_{i\alpha}$, $k_{i\alpha}$, and $r_{i\alpha}$, are related to the vocal fold tissue properties. The lumped-mass parameter values are presented in Table I.

With the solution $x_{i\alpha}$ and $v_{i\alpha}$ of the two-mass model, the glottal wall movement can be defined as follows: The glottal entry and exit correspond to $m_{1\alpha}$ and $m_{2\alpha}$ in a two-mass model, respectively. If the glottis is open, the displacement $X_{1\alpha}$ and the velocity $V_{1\alpha}$ of glottal entry are defined according to those of $m_{1\alpha}$ solved from the two-mass equation, i.e., $X_{1r} = x_{1r}$, $X_{1l} = -x_{1l}$, $V_{1r} = v_{1r}$ and $V_{1l} = -v_{1l}$. Similarly, the displacement $X_{2\alpha}$ and the velocity $V_{2\alpha}$ of glottal exit are set as $X_{2r} = x_{2r}$, $X_{2l} = -x_{2l}$, $V_{2r} = v_{2r}$, and $V_{2l} = -v_{2l}$. If m_{1l} and m_{1r} collide with each other, the displacements X_{1l} and X_{1r} of the glottal entry are set close to the midpoint between m_{1l} and m_{1r}, i.e., $X_{1l} = (x_{1r} - x_{1l} - x_g)/2$ and $X_{1r} = (x_{1r} - x_{1l} + x_g)/2$. Here, $x_g = 0.001$ cm represents a small space in order to avoid the failure of fluid solver due to zero grid volume during collision. The airflow velocity at the glottal entry is set as zero, which indicates that the airflow was interrupted at the collision point. If m_{2l} and m_{2r} collide with each other, the displacement and velocity at glottal exit follow definitions
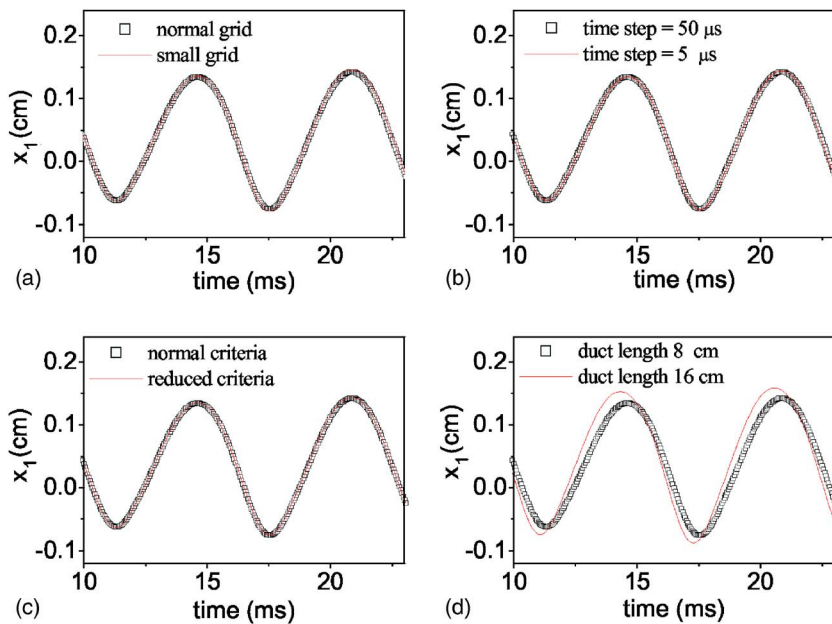
FIG. 2. (Color online) Verification of the composite model. The lung pressure is 2.0 kPa for this simulation. (a) Grid independence: The squares and the solid line represent the outputs of the model with the normal size grid used in the following simulation and the halved size grid, respectively. (b) Time step size independence: The squares and the solid line represent the output of the model solved with the time step 50 and 5 μs, respectively. (c) Convergence criteria independence: The squares represent the outputs of the model with the normal convergence criteria, which is used in the following simulation, and the solid line represents the model output with the convergence criteria reduced by a factor of 10. (d) Duct length: The squares and the solid line represent the model outputs with vocal tract length 8 and 16 cm, respectively.

similar to those at glottal entry, i. e., $X_{2l}=(x_{2r}-x_{2l}-x_g)/2$, $X_{2r}=(x_{2r}-x_{2l}+x_g)/2$. The displacement X_α and velocity V_α of the points on the glottal wall are linear interpolations of the displacement and velocity between the glottal entry and the glottal exit, that is, $X_\alpha(D_y)=X_{1\alpha}+(X_{2\alpha}-X_{1\alpha})\times D_y/D$ and $V_\alpha(D_y)=V_{1\alpha}+(V_{2\alpha}-V_{1\alpha})\times D_y/D$, where D is the length of glottis (= thickness of vocal fold) and D_y is the distance between this point and the glottal entry.

C. The self-oscillating solution

In our simulation, the NS equations and the two-mass vibration model were fully coupled and simultaneously solved in every iteration step. The glottal displacements ($x_{1\alpha}$ and $x_{2\alpha}$) and the glottal velocities ($v_{1\alpha}$ and $v_{2\alpha}$) solved from the two-mass model were applied on the glottal wall to define the boundary condition, including the glottal shape and the grid of the fluid domain (see Sec. II A); simultaneously, the airflow pressure solved from NS equations is applied on the two-mass model as the driving force.

Alternatively solving the NS equations and the two-mass equation, a self-oscillating solution was obtained from the iterative process. The entire airflow region was divided into 2600 grids with 2751 nodes with the first-order finite element scheme as shown in Fig. 1. The NS equations were solved by using the FLOTRAN CFD analysis of the finite-element software ANSYS. A custom program was developed in AN-SYS Parametric Design Language (APDL) to solve the two-mass equation with the fourth-order Runge-Kutta routine and manage the airflow-tissue interaction. In the time domain, both the NS equations and the two-mass equation were integrated with a time step of 50 μs. This composite model is different from a previous NS-based two-mass model (de Vries, 2002; 2003), where the glottal flow was assumed to be left-right symmetric and the discrete space grid refined at places of particular interest to represent the vocal fold movement. As such, asymmetric flow was not found in that model. In this model, a moving mesh based on arbitrary Lagrangian-Eulerian (ALE) formulation was used to deal with the glottal

movement. At every time step of the simulation, the new glottal wall position was calculated from the two-mass model. The surface displacement is defined by the boundary condition, while the position and size of the mesh inside the fluid flow region is modified to satisfy the boundary conditions at the moving interfaces. The offset of the node inside the flow region is inversely proportionate to its distance from the surface. The ALE formulation could provide a smooth spatial movement and therefore represent a better description of the vocal vibration. Furthermore, the left and the right glottis were both predicted in this model, which allowed us to investigate the asymmetry of flow in the glottis.

This composite model was verified to be independent of the grid size, time step, convergence criteria, and duct length (Luo and Pedley, 1995; 1996). Figure 2 presents the results. The model was first solved with a lung pressure of 2.0 kPa. Then, the models with the halved grid space, the small time step (5 μs), and the reduced convergence criteria were solved again with the same lung pressure. The solutions obtained are essentially identical as shown in Fig. 2(a)–2(c). The influence of the vocal tract was also examined by increasing the length of duct as shown in Fig. 2(d). It is found that the wave form of the model's output could be slightly changed because of different aerodynamic loading corresponding to different duct lengths. However, the asymmetry of airflow was not significantly altered by the duct length. These results suggested that the gird size, time step size, convergence criteria, and duct length used in this study are satisfactory.

III. RESULTS AND DISCUSSION

Elapsing 100 ms transient oscillation, we present the wave form of the sustained oscillation of the proposed composite model in Figs. 3(a), 3(c), and 3(e), where the lung pressure is 0.8 kPa (about 8 cm H_2O). For comparison, the corresponding output of the Bernoulli-based two-mass model (Steinecke and Herzel, 1995) with identical lumped-mass parameters is presented in Figs. 3(b), 3(d), and 3(f). In the

Tao *et al.*: Vocal fold vibration by asymmetric airflow   2273
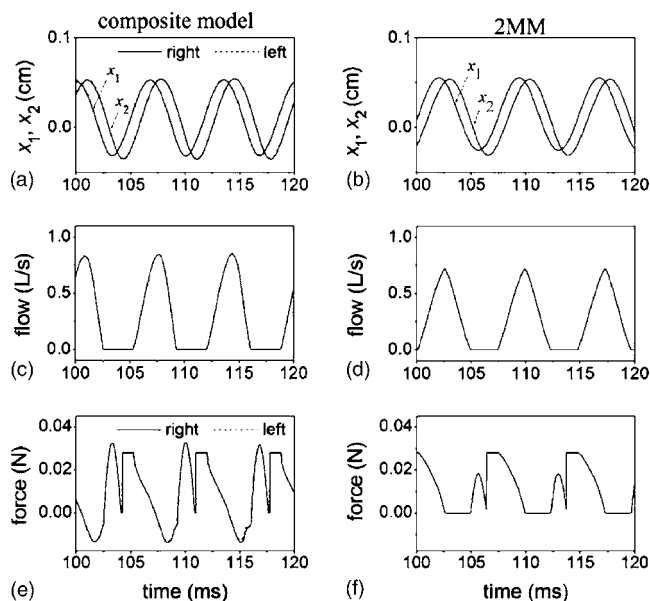
FIG. 3. Comparisons of the composite model and the Bernoulli-based two-mass model with lung a pressure of 0.8 kPa (about 8 cm H$_2$O). (a), (b) Displacement; (c), (d) volume velocity; and (e), (f) intraglottal pressure.

FIG. 4. Comparisons of the composite model and 2MM with a lung pressure of 2.0 kPa (about 20 cm H$_2$O). (a), (b) Displacement; (c), (d) volume velocity; and (e), (f) intraglottal pressure.

following discussion, we abbreviate the Bernoulli-based two-mass model as 2MM. Figure 3(a) shows the inferior and superior glottal displacement solved from the composite model. Figure 3(c) shows the wave form of the volume velocity of airflow in the glottis. It is seen that the composite model successfully produced a self-oscillating solution. The fundamental frequency predicted by the composite model was 146.5 Hz, which is higher than the fundamental frequency of 135.5 Hz predicted by the 2MM as shown in Figs. 3(b) and 3(d). In this study, the composite model and the 2MM have identical parameter values for the lumped mass as shown in Table I. The difference between the models is due to the different descriptions of the airflow in the glottis. For the 2MM, the airflow in the glottis is simplified to the Bernoulli relation. However, it is known that the Bernoulli assumption is valid only for steady, laminar, and incompressible fluid. For the composite model, NS equations are employed to describe the airflow field. The different airflow assumptions introduce differences in the intraglottal forces and further cause the differences in the vocal fold models' vibration. Figure 3(e) shows the intraglottal force predicted by the composite model, where the intraglottal force applied on the first mass is a composite force of the vocal fold collision and the airflow pressure. This force is due to airflow-tissue interaction. It is the driving force for the two-mass model and determines the vibratory characteristics of the vocal fold. It was found that there are two positive peaks and one negative peak during one cycle. The negative peak occurs during the precollision stage of one cycle. At that time, the airflow is accelerated in the narrow glottis and the Bernoulli effect generates the negative pressure in this stage. Following the negative peak, the higher positive peak is generated due to the strong vocal fold collision. The collision, rebuilt subglottal pressure, and the restoring force of the deformed vocal tissue push the vocal fold open. Then the air pressure is reestablished in the glottis and the second positive
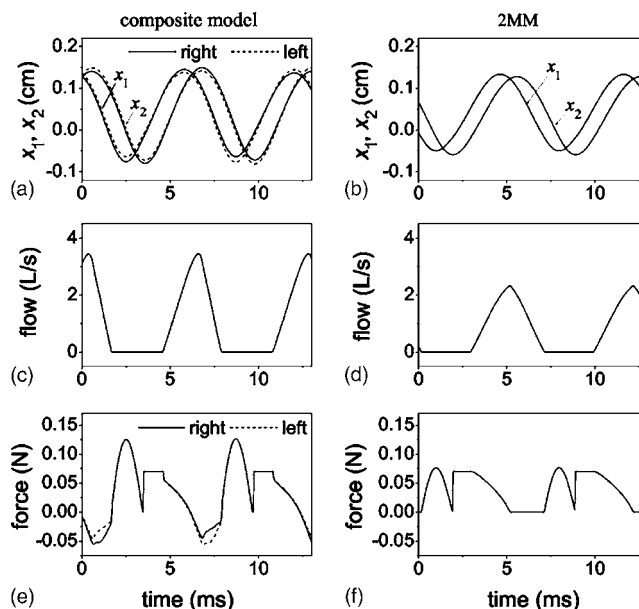
peak appears. These simulation results agree with experimental observations (Jiang and Titze, 1994) and previous modeling (Tao et al., 2006). The intraglottal force predicted by the 2MM is also presented in Fig. 3(f). It is seen that the intraglottal force in the 2MM is quite different from that in the composite model. The negative pressure is absent in the 2MM, and therefore, the vocal fold collision in the 2MM is also weaker than that predicted by the composite model. The above-mentioned results demonstrate that the composite model can generate a self-oscillating solution and that the NS equations in the composite model provide a better description of the airflow than the modified Bernoulli equations in the 2MM.

Under the low driving pressure, the two sides of the vocal folds show approximate symmetry; however, a high driving pressure increases the asymmetry of airflow and thus causes asymmetric vocal fold vibrations. Figure 4 compares the composite model and the 2MM under a lung pressure of 2.0 kPa (about 20 cm H$_2$O). It was found that the differences between the composite model and the 2MM are enlarged when the lung pressure is increased to 2.0 kPa. The fundamental frequency predicted by the composite model is 156.25 Hz, but for the 2MM it is 142.8 Hz. The flow velocity in the composite model is much higher than that in the 2MM. The negative pressure in the composite model becomes more significant, and therefore, the collision between two vocal folds is further enhanced. It can be seen that for the 2MM, the forces applied on the left and right vocal folds are always identical, however, in the composite model, the forces applied on the left and right vocal folds are visibly different, as shown in Fig. 4(e). Because the lumped mass parameters used in this model are exactly left-right symmetric, the above-presented result can only be explained by asymmetric flow in the glottis.

FIG. 5. (Color online) The velocity distribution during one vibratory cycle with lung pressure 0.8 kPa. (a) 6.00 ms; (b) 7.20 ms; (c) 8.40 ms; and (d) 8.85 ms.

To help understand the vocal fold vibration, it is beneficial to more thoroughly examine the airflow properties in the glottis during phonation. Figures 5 and 6 display the velocity distribution at four time instants during one vibratory cycle, where the lung pressure is 0.8 kPa for Fig. 5 and 2.0 kPa for Fig. 6. In addition, tissue-airflow interaction is an important factor in vocal fold vibration and voice production. The interaction between vocal fold and airflow mainly happens at the glottal wall. Therefore, Fig. 7 presents the pressure profiles on the glottal wall at four times during one vibratory cycle.

Figures 5(a) and 6(a) correspond to the phase of the glottis when it is just opening. Because the mucosal wave travels from the inferior side to the superior side, in a two-mass model there is a phase delay between the second mass representing the inferior side of the glottis and the first mass representing the superior side. The inferior side of the glottis is opened before the superior side; therefore, the glottis has a convergent shape at this time. The highest airflow velocity occurs at the exit of the glottis. However, the airflow velocity when the glottis is just opening is low for both the 0.8 and 2.0 kPa lung pressures. Moreover, the airflow velocity distribution in the glottis is left-right symmetric as shown in Figs. 5(a) and 6(a). Therefore, the air pressures applied on the left vocal folds are equal to those applied on the right vocal folds [see Fig. 7(a)].

Figures 5(b), 6(b), and 7(b) correspond to the phase when the glottis is fully opened. At this time, the glottal exit width is equal to the glottal entrance width. The glottal shape is rectangular. The volume velocity of airflow reaches its maximum value and the airflow velocity is also increased. For a lung pressure of 0.8 kPa, the velocity distribution in the glottis is still symmetric and the air pressures applied to the left and right vocal folds are identical. However, for a lung pressure of 2.0 kPa, the high velocity airflow does not remain symmetric. The airflow stream tends to stay attached to the left vocal fold, rather than follow a straight line. The

Coanda phenomenon occurs. Therefore, the velocity of the airflow near the left vocal fold is higher than that of airflow near the right vocal fold. The air pressure applied on the left vocal fold is lower than the pressure applied on the right vocal fold. Erath and Plesniak (2006a) found that the Coanda effect begins to occur when the acceleration of the forcing function was zero and the flow had reached maximum velocity. Therefore, they reasoned that the occurrence of the Coanda effect would mainly depend on the local unsteady acceleration rather than the cycle frequency. Our simulation shows that when the glottis is fully opened, the volume velocity of airflow reaches its maximum (Figs. 3 and 4) and the Coanda effect begins to occur [Fig. 5(b), 6(b), and 7(b)]. These results support Erath and Plesniak's conclusion.

Figures 5(c) and (d), 6(c) and (d), and Fig. 7(c) and (d) correspond to phases when the glottis is beginning to close. Because the glottal entrance is closed before the glottal exit, the glottis is divergent. The airflow velocity is very high in the glottis and the highest value appears at the glottal entrance. The divergent glottis causes the Coanda effect to be more significant. The airflow stream is significantly attached to one side of the vocal fold even when the lung pressure is only 0.8 kPa. The air pressure applied on the left and right vocal folds is significantly different for a lung pressure of 2.0 kPa. The left-right pressure difference is largest at the entrance to the glottis ($x=0.0$ cm). These results agree with previous static model simulations and experimental measurements (Pelorson et al., 1994; 1995; Pelorson and Hirschberg, 1997; Hofmans, 1998; Scherer et al., 2001; Hofmans et al., 2003; Shinwari et al., 2003) Erath and Plesniak (2006a) reported that the Coanda effect can occur for divergence angles of less than 20°. In our model, the maximum glottal divergent angles during one cycle are 7.7° at lung pressure 0.8 kPa and 19.5° at lung pressure 2.0 kPa. Therefore, the asymmetric airflow observed in this model all occurred for divergence angles less than 20°. These results concur with
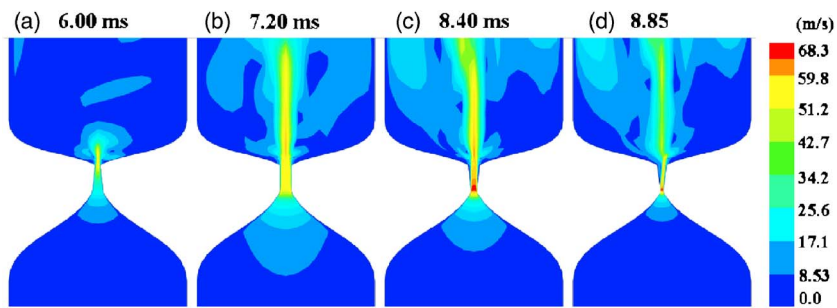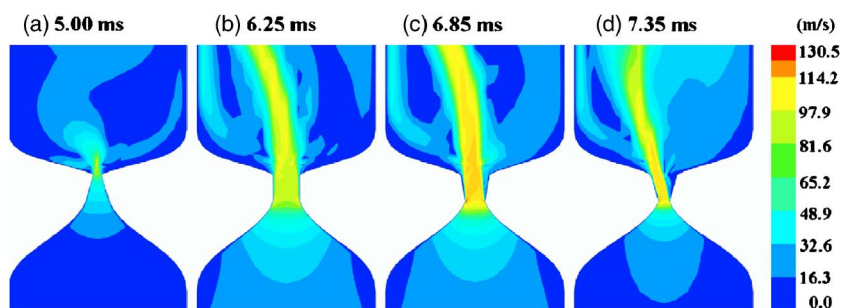

FIG. 6. (Color online) The velocity distribution during one vibratory cycle with lung pressure 2.0 kPa. (a) 5.00 ms; (b) 6.25 ms; (c) 6.85 ms; and (d) 7.35 ms.

Tao et al.: Vocal fold vibration by asymmetric airflow   2275
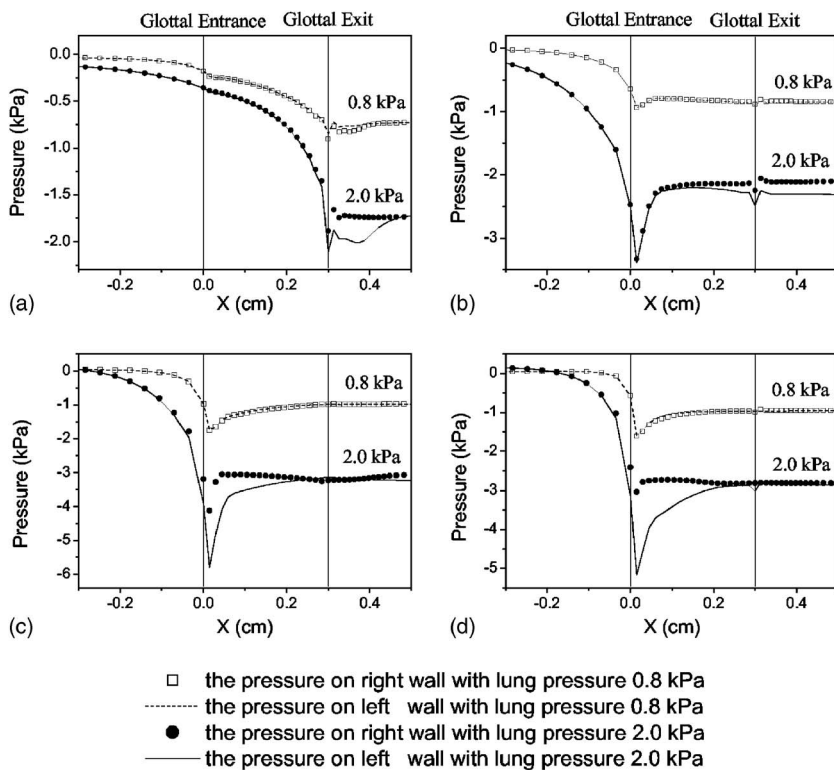
FIG. 7. Pressure drop from the trachea, where the times of (a)–(d) correspond to that of (a)–(d) in Figs. 5 and 6, respectively. (a) The convergent glottis at $t=6.0$ ms for lung pressure 0.8 kPa and $t=5.0$ ms for lung pressure 2.0 kPa. (b) The rectangular glottis at $t=7.20$ ms (0.8 kPa) and $t=6.25$ (2.0 kPa). (c) The divergent glottis at $t=8.4$ ms (0.8 kPa) $t=6.85$ ms (2.0 kPa). (d) The divergent glottis at $t=8.85$ and $t=7.35$ ms (2.0 kPa).

Erath and Plesniak's experiment. In addition, because the airflow velocity is very high (Reynolds number >4000) for lung pressure 2.0 kPa, the complex pattern can also be found downstream (Hofmans *et al.*, 2003; Iijima *et al.*, 1992)

The airflow asymmetry can be explained by the Coanda effect. Although the boundary condition is symmetric, the jet could be disturbed by many factors, including random vocal fold movement, asymmetric glottal passage (Erath and Plesniak, 2006a, b), turbulence due to high velocity flow (Hofmans *et al.*, 2003), and the asymmetric flow structure downstream (Neubauer *et al.*, 2007). In this study, the vocal fold part of this composite model is symmetric and does not include a random factor. Moreover, it has been verified that the airflow asymmetry observed in this model does not depend on the grid size, time step size, convergence criteria, or downstream duct length used in the simulation. The jet asymmetry is not artificially produced by a numerical instability but physically essential. Therefore, it would appear that what is most likely is that the asymmetric downstream flow patterns are disturbed and bring the jet close to one side of the vocal fold. Then, the fluid is accelerated along the proximate vocal fold surface, causing a pressure decrease in this surface. The pressure decrease pulls the airflow closer to the surface. The nearer the jet is to the surface, the greater the acceleration of fluid velocity and the greater decrease in pressure on the adjacent surface. Finally, the airflow becomes attached to one side of vocal fold (Erath and Plesniak, 2006b).

Using rapidly increasing volume flow through glottal models, Pelorson *et al.* (1995) and Hofmans *et al.* (1998; 2003) have found that the asymmetry of the flow due to the Coanda effect could happen in a divergent static glottis. However, they simultaneously found that it took a relatively long time to establish the Coanda effect. Therefore, it was considered that the asymmetry of flow could not occur during the production of normal voiced speech. However, using a computational and electrical analog model, Iijima *et al.* (1992) suggest that the establishment of the aerodynamics in the glottis is rapid (about 1 ms). Recently, Shinwari *et al.* (2003) also argued that the Coanda effect may have the best chance to form during phonation because the glottis begins to take on a divergent shape only near the peak of the glottal flow. Therefore, flow asymmetry remains an arguable question because of the lack of direct evidence for the presence and influence of the Coanda effect during phonation (in a vibratory glottis). The composite glottal model in this experiment could simultaneously predict the changing glottal shape due to vocal fold vibration and the aerodynamics in the glottis. The simulation results show that the airflow is asymmetric due to the Coanda effect. Coanda effects cause the air pressure applied on the right vocal fold to be different from that applied on left vocal fold. The results are numerical evidence for the presence of asymmetric flow during phonation.

The composite model not only has the capability to describe the aerodynamics in a vibratory glottis, but it also allows researchers to study vocal fold vibration under the driving of asymmetric airflow in the glottis. Figure 8 compares the left and right vocal fold vibrations under lung pressures of 0.8 and 2.0 kPa, respectively. If the left vocal fold and the right vocal fold vibrations are symmetric about the midline ($x=0$), all tracks should concentrate at the diagonal of the $x_{1l}-x_{1r}$ phase space, otherwise, the track could depart from the diagonal, meaning that the vocal fold vibrations are not left-right symmetric. Figure 8(a) shows that the track slightly departs from the diagonal and that the vocal fold vibration is asymmetric about the midline ($x=0$). If the lung
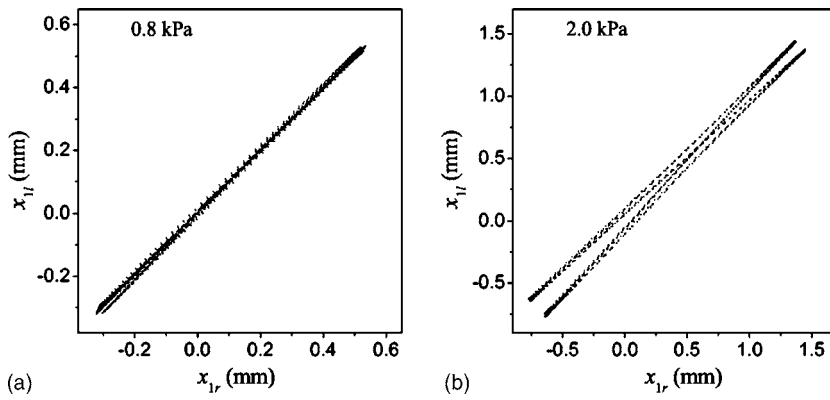
FIG. 8. Asymmetric vocal fold vibration due to the asymmetric pressure flow distribution in the glottis. (a) Lung pressure is 0.8 kPa. (b) Lung pressure is 2.0 kPa.

pressure is increased to 2.0 kPa, then the asymmetry becomes more significant [see Fig. 8(b)]. It is noticed that the system parameters of the two-mass model in this composite model are completely left-right symmetric. The asymmetric vibration is induced purely by the asymmetric driving force due to the Coanda effect. Because the flow asymmetry under a lung pressure of 2.0 kPa is more serious than that under a lung pressure of 0.8 kPa, the vibratory asymmetry under 2.0 kPa is also larger than that under 0.8 kPa.

The ratio between the standard deviation of x_{1r} and the standard deviation of $(x_{1l}-x_{1r})$ was defined to quantitatively describe the asymmetric vocal fold vibration due to the Coanda effect:

$$\text{Asymmetry} = \frac{\langle x_{1r}\rangle \sum_{n}\left[(x_{1l}-x_{1r})-\langle x_{1l}-x_{1r}\rangle\right]}{\langle x_{1l}-x_{1r}\rangle \sum_{n}\left[x_{1r}-\langle x_{1r}\rangle\right]} \times 100\% ,$$

(6)

where $\langle x(n)\rangle$ represents the mean value of $x(n)$ with respect to time and n is the discrete time. Figure 9(a) shows $\text{SD}(x_{1r})$ and $\text{SD}(x_{1l}-x_{1r})$ as a function of lung pressure, where $\text{SD}(x)$ denotes the standard deviation of x. It was found that both $\text{SD}(x_{1r})$ and $\text{SD}(x_{1l}-x_{1r})$ are increased by increasing lung pressure. The $\text{SD}(x_{1r})$ results can be fitted very well by the linear line $[y(x)=a(x-b)$ with $a=0.03567$ and $b=0.00554]$. However, $\text{SD}(x_{1l}-x_{1r})$ is best fit by the quadratic curve $[y(x)=a+bx+cx^2$ with $a=0.00205$, $b=-0.004987$ and $c=0.00401]$ within the lung pressure range in this study. Therefore, the asymmetry of the vocal fold vibration is also increased from 1.87% to 11.2% when the lung pressure is increased from 0.6 to 2.0 kPa. These results suggest that the asymmetry of airflow in the glottis could be an important factor influencing the dynamic behavior of vocal fold vibration.

IV. CONCLUSION

This study proposed a composite glottal model by combining together the Navier-Stokes equations and a two-mass vocal fold model. Numerical simulation showed that this model can generate a self-oscillating solution. Therefore, it has the ability to predict both the behaviors of glottal movement and aerodynamics in a vibratory glottis. Compared to the Bernoulli-based two-mass model, it was found that this composite model provides a better description of the glottal airflow. The complex asymmetric airflow stream during phonation was predicted by this model. The glottis had a divergent shape before the glottis is closed and the glottal airflow was asymmetric due to Coanda effects. The airflow stream tended to stay attached to one side of the vocal fold, rather than following a straight line in its original direction. The asymmetric airflow distribution further induced asymmetric pressure on the left and right vocal folds. This numerical simulation indicated that the asymmetric flow due to Coanda effects could exist in a vibratory glottis during phonation. It was found that even when the left vocal fold and right vocal fold have identical tissue properties, the vocal fold vibrations could still be asymmetric due to the asymmetric airflow. Moreover, the vibration asymmetry is increased with increasing lung pressure within the pressure range (0.6–2.0 kPa) applied in this study.
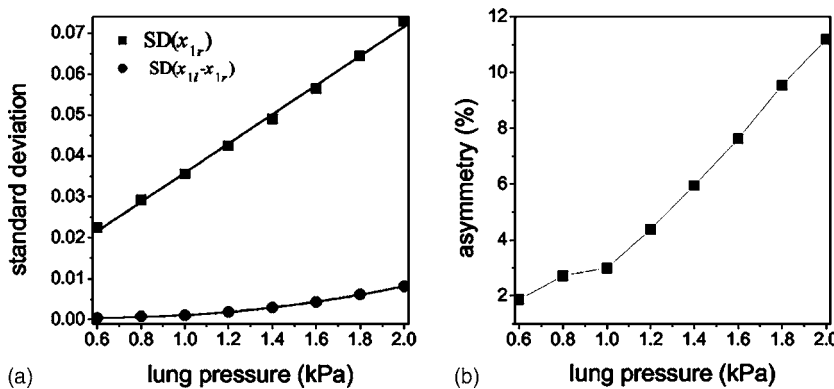
FIG. 9. The influence of lung pressure on vocal fold vibration. (a) The standard deviation of x_{1r} and $(x_{1l}-x_{1r})$ as a function of lung pressure. (b) The vibration asymmetry as a function of lung pressure.

In summary, this study showed that besides asymmetries in vocal fold geometrical dimensions and tissue properties, glottal aerodynamics are another important factor inducing asymmetric vocal fold vibration. Knowledge of glottal aerodynamics is an important part of a more complete understanding of vocal fold vibrations and phonation. Our composite model simulation showed a qualitative consistency with previous experimental observations, suggesting that this composite model could be a useful tool in studying the aerodynamic characteristics of phonation. The vocal tract acoustics are ignored in the proposed composite model. However, it is known that vocal tract acoustics could also be an important factor affecting vocal fold vibration (Titze, 1988; Laje *et al.*, 2001). Therefore, further improvement upon this model is still important.

ACKNOWLEDGMENT

Alipour, F., Scherer, R., and Knowles, J. (**1996**). "Velocity distributions in glottal models," J. Voice **10**, 50–58.

Alipour, F., and Scherer, R. C. (**2002**). "Pressure and velocity profiles in a static mechanical hemilarynx model," J. Acoust. Soc. Am. **112**, 2996–3003.

Alipour, F., and Scherer, R. C. (**2004**). "Flow separation in a computational oscillating vocal fold model," J. Acoust. Soc. Am. **116**, 1710–1719.

Alipour, F., and Titze, I. R. (**1996**). "Combined simulation of two-dimensional airflow and vocal fold vibration," in *Vocal Physiology: Controlling Complexity and Chaos*, Edited by P. J. Davis and N. H. Fletcher (Singular, San Diego, 1996). pp. 17–29.

De Oliveira Rosa, M., Pereira, J. C., Grellet, M., and Alwan, A. (**2003**). "A contribution to simulating a three-dimensional larynx model using the finite element method," J. Acoust. Soc. Am. **114**, 2893–2905.

de Vries, M. P. Hamburg, M. C., Schutte, H. K., Verkerke, G. J., and Veldman A. E. P. (**2003**). "Numerical simulation of self-sustained oscillation of a voice-producing element based on Navier-Stokes equations and the finite element method," J. Acoust. Soc. Am. **113**, 2077–2083.

de Vries, M. P., Schutte, H. K., Veldman, A. E. P., and Verkerke, G. J. (**2002**). "Glottal flow through a two-mass model: Comparison of Navier-Stokes solutions with simplified models," J. Acoust. Soc. Am. **111**, 1847–1853.

Erath, B. D., and Plesniak, M. W. (**2006a**). "The occurrence of the Coanda effect in pulsatile flow through static models of the human vocal folds," J. Acoust. Soc. Am. **120** 1000–1011.

Erath, B. D., and Plesniak, M. W. (**2006b**). "An investigation of jet trajectory in flow through scaled vocal fold models with asymmetric glottal passages," Exp. Fluids **41**, 735–748.

Erath, B. D., and Plesniak, M. W. (**2006c**). "An investigation of bimodal jet trajectory in flow through scaled models of the human vocal tract," Exp. Fluids **40**, 683–696.

Guo, C.-G., and Scherer, R. C. (**1993**). "Finite element simulation of glottal flow and pressure," J. Acoust. Soc. Am. **94**, 688–700.

Hofmans, G. C. J. (**1998**). *Vortex Sound in Confined Flows* (Technische, Universiteit Eindhoven, Eindhoven).

Hofmans, G. C., 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.

Iijima, H., Miki, N., and Nagai, N. (**1992**). "Glottal impedance based on a finite element analysis of two-dimensional unsteady viscous flow in a static glottis," IEEE Trans. Signal Process., **40**, 2125–2135.

Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J. **51**, 1233–1268.

Jiang, J. J., and Titze, I. R. (**1994**). "Measurement of vocal fold intraglottal pressure and impact stress," J. Voice **8**, 132–144.

Jiang, J. J., Zhang, Y., and Stern, J. (**2001**). "Modeling of chaotic vibrations in symmetric vocal folds," J. Acoust. Soc. Am. **110**, 2120–2128.

Kaiser, J. F. (**1983**). "Some observations on vocal tract operation from a fluid flow point of view," in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, edited by I. R. Titze and R. S. Scherer (The Denver Center for Performing Arts, Denver, CO), pp. 358–386.

Kucinschi, B. R., Scherer, R. C., DeWitt, K. J., and Ng, T. T. M. (**2006**). "An experimental analysis of the pressures and flows within a driven mechanical model of phonation," J. Acoust. Soc. Am. **119**, 3011–3021.

Laje, R., Gardner, T., and Mindlin, G. B. (**2001**). "Continuous model for vocal fold oscillations to study the effect of feedback," Phys. Rev. E, **64**, 056201.

Lucero, J. C. (**1996**). "Relation between the phonation threshold pressure and the prephonatory glottal width in a rectangular glottis," J. Acoust. Soc. Am. **100**, 2551–2554.

Luo, X. Y., and Pedley, T. J. (**1995**). "A numerical simulation of steady flow in a 2-D collapsible channel," J. Fluids Struct. **9**, 149–174.

Luo, X. Y., and Pedley, T. J. (**1996**). "A numerical simulation of unsteady flow in a two dimensional collapsible channel," J. Fluid Mech. **314**, 191–225.

Neubauer, J., Zhang, Z., Miraghaie, R., and Berry, D. A. (**2007**). "Coherent structures of the near field flow in a self-oscillating physical model of the vocal folds," J. Acoust. Soc. Am. **121**, 1102–1118.

Pelorson, X., Hirschberg, A., Van Hassel, R. R., and Wijnands, A. P. J. (**1994**). "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model," J. Acoust. Soc. Am. **96**, 3416–3431.

Pelorson, X., Hirschberg, A., Wijnands, A. P. J., and Bailliet, H. (**1995**). "Description of the flow through in-vitro models of the glottis during phonation," Acta Acust. **3**, 191–202.

Pelorson, X., and Hirschberg, A. (**1997**). "In vitro study of the glottal and supra-glottal flow during phonation," in *Proceedings of the Fourth Congress on Acoustics*, Marseille, Teknea, Toulouse, pp. 337–340.

Scherer, R. C., Shinwari, D., De Witt, K. J. Zhang, C., Kucinschi, B. R., and Afjeh, A. A. (**2001**). "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degree," J. Acoust. Soc. Am. **109**, 1616–1630.

Scherer, R. C., Shinwari, D., De Witt, K. J. Zhang, C., Kucinschi, B. R., Afjeh, A. A. (**2002**). "Intraglottal pressure distributions for a symmetric and oblique glottis with a uniform duct," J. Acoust. Soc. Am. **112**, 1253–1256.

Shinwari, D., Scherer, R. C., DeWitt, 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.

Steinecke, I., and Herzel, H. (**1995**). "Bifurcations in a vocal fold model," Nonlinear Dyn. **7**, 53–64.

Tao, C., and Jiang, J. J. (**2006**). "Anterior-posterior biphonation in a finite element model of vocal fold vibration," J. Acoust. Soc. Am. **120**, 1570–1577.

Tao, C., Jiang, J. J., and Zhang, Y. (**2006**). "Simulation of vocal fold impact pressures with a self-oscillating finite-element model," J. Acoust. Soc. Am. **119**, 3987–3994.

Teager, H. M., and Teager, S. M. (**1983**). "Active fluid dynamics voice production, or there is a unicorn in the garden," in *Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control*, edited by I. R. Titze and R. S. Scherer (The Denver Center for Performing Arts, Denver, CO), pp. 387–394.

Thomson, S. L., Mongeau, L., and Frankel, S. H. (**2005**). "Aerodynamic transfer of energy to the vocal folds," J. Acoust. Soc. Am. **118**, 1689–1701.

Titze, I. R. (**1988**). "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**, 1536–1552.

Titze, I. R. (**1994**). *Principles of Voice Production*. (Prentice Hall, Upper Saddle River, NJ).

Zhang, C., Zhao, W., Frankel, S. H., and Mongeau, L. (**2002**). "Computational aeroacoustics of phonation, Part II. Effects of flow parameters and ventricular folds," J. Acoust. Soc. Am. **112**, 2147–2154.

Zhao, W., Zhang, C., Frankel, S. H., and Mongeau, L. (**2002**). "Computational aeroacoustics of phonation, Part I. Computational methods and sound generation mechanisms," J. Acoust. Soc. Am. **112**, 2134–2146.

Physical mechanisms of phonation onset: A linear stability analysis of an aeroelastic continuum model of phonation

Zhaoyan Zhang,[a] Juergen Neubauer, and David A. Berry
The Laryngeal Dynamics Laboratory, Division of Head and Neck Surgey, David Geffen School of Medicine at UCLA, 31-24 Rehabilitation Center, 1000 Veteran Ave., Los Angeles, California 90095-1794

In an investigation of phonation onset, a linear stability analysis was performed on a two-dimensional, aeroelastic, continuum model of phonation. The model consisted of a vocal fold-shaped constriction situated in a rigid pipe coupled to a potential flow which separated at the superior edge of the vocal fold. The vocal fold constriction was modeled as a plane-strain linear elastic layer. The dominant eigenvalues and eigenmodes of the fluid-structure-interaction system were investigated as a function of glottal airflow. To investigate specific aerodynamic mechanisms of phonation onset, individual components of the glottal airflow (e.g., flow-induced stiffness, inertia, and damping) were systematically added to the driving force. The investigations suggested that flow-induced stiffness was the primary mechanism of phonation onset, involving the synchronization of two structural eigenmodes. Only under conditions of negligible structural damping and a restricted set of vocal fold geometries did flow-induced damping become the primary mechanism of phonation onset. However, for moderate to high structural damping and a more generalized set of vocal fold geometries, flow-induced stiffness remained the primary mechanism of phonation onset. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2773949]

I. INTRODUCTION

Self-sustained oscillation, or flutter, is a general phenomenon which commonly occurs in coupled fluid-structure systems. Vocal fold vibration is one example of this phenomenon. Other examples can be found in biological systems (e.g., the flow-induced instabilities of the blood vessels in arterial stenoses, bronchial airway oscillations during snoring, see Grotberg and Jensen, 2004) and mechanical systems (e.g., airfoil flutter, see Dowell and Hall, 2001; brass instruments, Cullen *et al.*, 2000). There have been extensive investigations on the interaction of external flows with compliant plates or membranes (e.g., Benjamin, 1960; Landahl, 1962; Benjamin, 1963; Pierucci, 1977; Carpenter and Garrad, 1985, 1986; Duncan, 1987; Yeo, 1988, 1992; Lucey and Carpenter, 1993; Yeo *et al.*, 1994). Fluid-structure instabilities due to internal flows through compliant tubes have also been investigated (e.g., Holmes, 1977; Matsuzaki and Fung, 1977, 1979; Luo and Pedley, 1996; Huang, 2001). Although instabilities in coupled fluid-structure systems can originate from either the fluid or the structure alone, previous investigations have shown that flutter also arises from the coalescence of two eigenmodes (1:1 entrainment) to form a coupled-mode flutter. These two interacting eigenmodes could be either a flow mode coupled to a structural mode (e.g., the vortex-induced vibration, see Williamson and Govardhan, 2004), or two structural modes (e.g., Auregan and Depollier, 1995; Cullen *et al.*, 2000).

During normal phonation, it is generally understood that vocal fold vibration involves the excitation of two or more eigenmodes of the vocal system. For example, early research with the one-mass model (Flanagan and Landgraf, 1968) showed that self-oscillation of the folds could not occur unless coupled with an acoustic resonance of the vocal tract. However, without a vocal tract, the two-mass model (Ishizaka and Flanagan, 1972) was able to vibrate through the synchronization of two structural eigenmodes, one with the two masses vibrating in phase with each other, and the other with the two masses vibrating out of phase with each other. The out-of-phase motion created an alternating divergent and convergent glottis, regulating the intraglottal air pressure. The in-phase motion opened and closed the glottal channel, modulating the airflow. The synchronization of empirical modes of vocal fold vibration has also been observed in both theoretical and laboratory models of phonation (Berry *et al.*, 1994, 2001; Zhang *et al.*, 2006b).

Titze (1988) argued that the propagation of the mucosal wave, or the synchronization of two eigenmodes, facilitated self-oscillation by providing a sufficient, net transfer of energy from the airflow (fluid) to the tissue (structure) to overcome structural damping. However, the linear stability analysis completed by Titze (1988) did not compute the underlying eigenmodes of the two-dimensional vocal fold continuum, but assumed the eigenmodes *a priori*. Also, the temporal phase relationship between the two eigenmodes was imposed, rather than treated as a dynamic variable of the system. Thus, the mechanisms of phonation onset in terms of eigenmode synchronization and the role of glottal aerodynamics in the synchronization process could not be studied.

Using the two-mass model, Ishizaka and colleagues performed a linear stability analysis to investigate physical mechanisms of phonation onset (Ishizaka and Matsudaira, 1972; Ishizaka, 1981, 1988). In these investigations, they

---

modeled the aerodynamic driving force as a flow-induced stiffness term and showed that the two eigenfrequencies of the two-mass model approximated each other as the flow-induced stiffness was increased (e.g., see pp. 52, 53 of Ishizaka and Matsudaira, 1972; pp. 233, 238 of Ishizaka, 1981). At a threshold of the flow-induced stiffness or subglottal pressure, the two eigenfrequencies synchronized (1:1 entrainment), resulting in phonation onset. Although these studies provided new insight into physical mechanisms of phonation onset, a direct translation of these findings to practical applications has been problematic, perhaps in part because the key model parameters, such as masses and stiffnesses of the coupled oscillators, were immeasurable and difficult to relate to the anatomical structure of the vocal folds (Titze, 1988). Indeed, empirical measurement of the geometric and viscoelastic properties of vocal fold tissues may be more naturally and directly related to the physical properties of continuum models of vocal fold vibration (e.g., Titze *et al.*, 1995; Deverge *et al.*, 2003; Cook and Mongeau, 2007), as opposed to lumped-element models (e.g., Ishizaka and Flanagan, 1972; Steinecke and Herzel, 1995; Story and Titze, 1995; Horacek and Svec, 2002). Furthermore, several investigations have suggested that the eigenfrequency structure of a vocal fold continuum (Berry and Titze, 1996; Cook and Mongeau, 2007) may be much different from the eigenfrequency structure of a lumped-element model (Ishizaka and Matsudaira, 1972; Ishizaka and Flanagan, 1972; Ishizaka, 1981, 1988) resulting in distinct physical mechanisms of phonation onset (Zhang *et al.*, 2006b).

The motivation of the present study is to investigate physical mechanisms of phonation onset in a two-dimensional aeroelastic continuum model of the vocal folds, similar to the investigations of Ishizaka and colleagues for the two-mass model (Ishizaka and Matsudaira, 1972; Ishizaka, 1981, 1988). Similar to Titze (1988), this investigation will apply linear stability analysis to a two-dimensional continuum model of the vocal folds. However, rather than prescribing traveling wave motion and assuming a fixed eigenmode entrainment *a priori*, the present study will utilize a self-oscillating continuum model to investigate the role of glottal aerodynamics in initiating phonation. Using linear plane-strain theory, the vocal fold tissues will be modeled as a two-dimensional continuum, which will be driven by a one-dimensional potential glottal flow with a fixed flow separation location. However, the potential flow utilized here will be more generalized than the one utilized by Ishizaka and colleagues (Ishizaka and Matsudaira, 1972; Ishizaka, 1981), including components not only for flow-induced stiffness, but also flow-induced damping, and flow-induced inertia. Furthermore, in order to investigate specific, aerodynamic mechanisms of phonation onset, the individual components of the glottal airflow will be systematically and separately added to the driving force. That is, initially, only flow-induced stiffness will be considered, as per previous investigations by Ishizaka and colleagues (Ishizaka and Matsudaira, 1972; Ishizaka, 1981, 1988). Next, flow-induced inertia will be added to the driving force, followed by flow-induced damping.
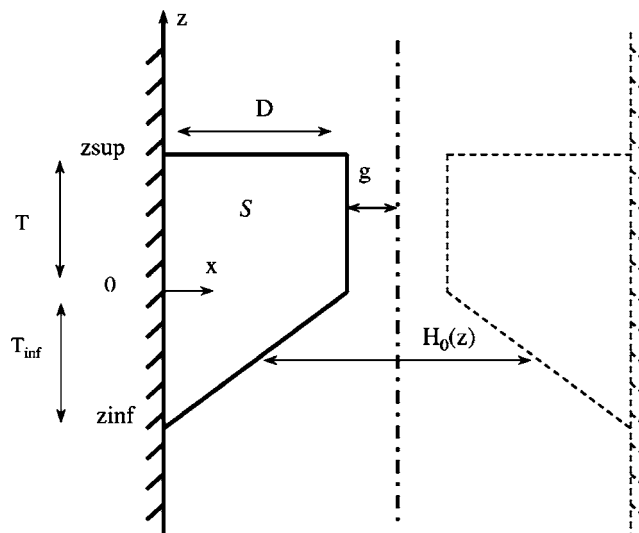
FIG. 1. The two-dimensional vocal fold model and the glottal channel. The flow direction is along the positive z axis. The coupled vocal folds-flow system was assumed to be symmetric about the glottal channel centerline so that only the left half of the system was considered in this study. D is the vocal fold depth in the medial-lateral direction; $T = z\mathrm{sup}$ and $T_{\mathrm{inf}} = -z\mathrm{inf}$ are the thicknesses of the vocal fold in the superior ($z > 0$) and inferior ($z < 0$) sections in the flow direction, respectively; H_0 is the prephonatory glottal channel width; g is the prephonatory glottal half width in the superior part ($z > 0$) of the glottal channel; $z\mathrm{sup}$ and $z\mathrm{inf}$ are the vertical coordinates of the outlet and inlet of the glottal channel, respectively.

II. MODEL DESCRIPTION

We considered a one-dimensional potential flow through a constriction formed by two symmetric, two-dimensional, isotropic vocal fold-shaped elastic layers. For simplicity, left-right symmetry of the flow and vocal fold vibration about the glottal channel centerline was imposed and therefore only half the system was considered (Fig. 1). A nondimensional formulation was used in this study. The vocal fold thickness T, vocal fold density ρ_{vf}, and the wave velocity of the vocal fold structure $\sqrt{E/\rho_{vf}}$ were used as the reference length, density, and velocity scales, respectively. All variables in the following are in nondimensional forms.

A. Structural model

The vocal folds were modeled as a two-dimensional, isotropic, plane-strain, elastic layer, as shown in Fig. 1. As mentioned above, only one vocal fold was considered. The lateral surface of the vocal fold was fixed to the pipe wall, while the other three sides were fluid-structure-interaction (FSI) boundaries. A linear stress-strain relationship was assumed for the vocal fold material. For a displacement field of the vocal fold structure $[\xi, \eta]$ (the x and z components of the vocal fold displacement, respectively), the corresponding strains and stresses are:

$$\varepsilon = [\varepsilon_x, \varepsilon_z, \gamma_{xz}] = \left[\frac{\partial \xi}{\partial x}, \frac{\partial \eta}{\partial z}, \frac{\partial \xi}{\partial z} + \frac{\partial \eta}{\partial x} \right], \quad (1)$$

$$\tau = [\tau_x, \tau_z, \tau_{xz}] = G\varepsilon, \quad (2)$$

where the stress-stain material matrix G is

$$G = \frac{E(1-\nu)}{(1+\nu)(1-2\nu)} \begin{bmatrix} 1 & \dfrac{\nu}{1-\nu} & 0 \\[2ex] \dfrac{\nu}{1-\nu} & 1 & 0 \\[2ex] 0 & 0 & \dfrac{1-2\nu}{2(1-\nu)} \end{bmatrix}, \qquad (3)$$

where E and ν are the Young's modulus and the Poisson's ratio of the vocal fold material, respectively. The associated kinetic energy V and potential energy U of the vocal fold structure are

$$V = \frac{1}{2} \int \int_S (\dot{\xi}^2 + \dot{\eta}^2) \rho_{vf} dS; \qquad (4)$$

$$U = \frac{1}{2} \int \int_S (\varepsilon_x \tau_x + \varepsilon_z \tau_z + \gamma_{xz} \tau_{xz}) dS,$$

where S is the two-dimensional vocal fold volume, and ρ_{vf} is the vocal fold density.

The governing equations of motion for the vocal fold can be obtained by substituting the above expressions into the Lagrange's equation

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_r}\right) - \frac{\partial L}{\partial q_r} = Q_r, \quad r = 1, 2, \ldots, N, \qquad (5)$$

where the Lagrangian $L = V - U$, and Q_r is the generalized force acting on the rth generalized coordinate q_r. The generalized coordinates will be defined in Sec. II C.

In Eq. (4), the contribution of the mean stress and deformation fields of the vocal fold due to the mean flow field is not included as part of the potential energy. However, it can be shown that, for a linear stress-strain relationship of the vocal fold material, this contribution is balanced by the mean flow pressure applied to the vocal fold along the interface. Accordingly, the generalized force in our formulation includes only the contribution from the fluctuating flow pressure.

Equation (5) can be reorganized and written in matrix form as

$$M\ddot{q} + Kq = Q, \qquad (6)$$

where the mass and stiffness matrices M and K are defined as

$$M_{ij} = \frac{\partial}{\partial \ddot{q}_j}\left(\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right)\right); \quad K_{ij} = \frac{\partial}{\partial q_j}\left(-\frac{\partial L}{\partial q_i}\right). \qquad (7)$$

The corresponding generalized force Q is

$$Q_r = -\int_{\ell_{\text{FSI}}} \left(p \frac{\partial \xi}{\partial q_r} n_x + p \frac{\partial \eta}{\partial q_r} n_z\right) d\ell, \quad r = 1, 2, \ldots, N, \qquad (8)$$

where ℓ_{FSI} denotes the fluid-structure interface with its normal vector n pointing outward from the vocal folds, and p is the flow pressure on the fluid-structure interface. The minus sign in the equation ensures that the force applied by the flow pressure on the vocal fold surface is always directed into the vocal fold structure. To solve the governing equations, an additional equation relating the pressure p and the vocal fold

displacement field $[\xi, \eta]$ is needed, which will be derived in Sec. II B.

A proportional structural damping was assumed for the vocal fold material. Equation (6) can be modified to include structural damping

$$M\ddot{q} + C\dot{q} + Kq = Q, \qquad (9)$$

where C is the structural damping matrix. In this study, a constant loss factor σ was used, which relates the mass and damping matrices as follows:

$$C = \sigma \omega M, \qquad (10)$$

where ω is the angular frequency.

B. Flow model

The following assumptions were made to simplify the flow through the glottis:

1. The flow was assumed to be a one-dimensional, incompressible, potential flow up to the point of flow separation.
2. The flow separation location was fixed at the superior edge of the vocal folds $z = zsup$ (Fig. 1). The jet flow discharged into open space with zero pressure recovery ($P_{0,zsup} = 0$).
3. The flow rate at the vocal fold inlet was constant.

In this study, the mean jet velocity at the flow separation point was used as the major model control parameter. For a given mean jet velocity U_j, the mean velocity distribution U_0 and mean pressure P_0 along the glottal channel were

$$U_0(z) = \frac{2gU_j}{H_0(z)}; \quad P_0(z) = \frac{1}{2}\rho_f U_j^2\left(1 - \frac{4g^2}{H_0^2}\right), \qquad (11)$$

where ρ_f is the air density, g is the mean glottal half width in the superior section ($z > 0$) of the glottal channel (Fig. 1), and H_0 is the mean width of the glottal channel.

The governing equations for the fluctuating flow velocity u and pressure p were obtained by linearization of the one-dimensional continuity equation and Bernoulli's equation around the mean state (U_0, P_0, H_0) [see, e.g., Lighthill, 1978]

$$\frac{\partial h}{\partial t} + \frac{\partial(U_0 h)}{\partial z} + \frac{\partial(u H_0)}{\partial z} = 0, \qquad (12)$$

$$\frac{\partial u}{\partial t} + U_0 \frac{\partial u}{\partial z} + u \frac{\partial U_0}{\partial z} + \frac{1}{\rho_f} \frac{\partial p}{\partial z} = 0, \qquad (13)$$

where h is the fluctuating component of the glottal channel width. The fluctuating width of the glottal channel is related to the displacement of the vocal fold structure at the fluid-structure interface as

$$h = -2\xi|_{\ell_{\text{FSI}}}. \qquad (14)$$

The factor of 2 appears due to the left-right symmetry of the vocal fold vibration. The boundary conditions for the flow are, according to assumptions 2 and 3:

$$u_{z=z\text{inf}} = 0, \quad p_{z=z\text{sup}} = 0, \tag{15}$$

where $z\text{inf}$ and $z\text{sup}$ are the vertical coordinates of the inlet and outlet of the glottal channel (Fig. 1). Other types of boundary conditions can be also used to model the coupling of the sub- and supraglottal acoustics and the vocal fold vibration. This will be discussed in future work.

The flow pressure within the glottal channel can be obtained by first integrating the linearized continuity Eq. (12), and then the linearized momentum Eq. (13) along the fluid-structure interface. After applying the boundary conditions in Eq. (15), this yields

$$\frac{p(z,t)}{\rho_f} = \int_{z\text{sup}}^{z} \left(\frac{\partial}{\partial t} + U_0 \frac{\partial}{\partial z} + \frac{\partial U_0}{\partial z} \right)$$
$$\times \left[\frac{1}{H_0} \int_{z\text{inf}}^{z} \left(\frac{\partial}{\partial t} + U_0 \frac{\partial}{\partial z} + \frac{\partial U_0}{\partial z} \right) h dz \right] dz$$
$$= p_0(h) + p_1(h_t) + p_2(h_{tt}), \tag{16}$$

where

$$p_0(h) = \int_{z\text{sup}}^{z} \left(U_0 \frac{\partial}{\partial z} + \frac{\partial U_0}{\partial z} \right) \left[\frac{1}{H_0} \int_{z\text{inf}}^{z} \left(U_0 \frac{\partial}{\partial z} \right. \right.$$
$$\left. \left. + \frac{\partial U_0}{\partial z} \right) h dz \right] dz = \frac{U_0^2 h}{H_0} - \left(\frac{U_0^2 h}{H_0} \right) \bigg|_{z\text{sup}}$$

$$p_1(h_t) = \int_{z\text{sup}}^{z} \left[\frac{1}{H_0} \int_{z\text{inf}}^{z} \left(U_0 \frac{\partial}{\partial z} + \frac{\partial U_0}{\partial z} \right) h_t dz \right] dz$$
$$+ \int_{z\text{sup}}^{z} \left(U_0 \frac{\partial}{\partial z} + \frac{\partial U_0}{\partial z} \right) \left[\frac{1}{H_0} \int_{z\text{inf}}^{z} h_t dz \right] dz, \tag{17}$$

$$p_2(h_{tt}) = \int_{z\text{sup}}^{z} \frac{1}{H_0} \int_{z\text{inf}}^{z} h_{tt} dz dz$$

where the subscripts t and tt denote first and second derivative with respect to time, respectively. The generalized force Q can be then obtained by substitution of Eq. (16) into Eq. (8), and regrouping in the same form as Eq. (16)

$$Q = Q_0 q + Q_1 \dot{q} + Q_2 \ddot{q}, \tag{18}$$

where Q_0, Q_1, and Q_2 represent the flow-induced stiffness, flow-induced damping, and flow-induced inertia, respectively.

Equations (16) and (17) show that the fluctuating flow pressure on the FSI surface is composed of three parts, which represent the flow-induced stiffness, damping, and inertia, respectively. The first term, scaling with $\rho_f U_0^2 / H_0$, represents the flow-induced stiffness. This term is generally the most dominant term of the three, especially at large flow velocities. Note that the matrix Q_0 is asymmetric, mainly due to the imposed flow separation at the superior edge (the second term in the equation for $p_0(h)$ which is nonzero). We will show this feature of the flow-induced stiffness matrix to be an essential component of phonation onset. The second term in Eq. (16) is a first-order term in time, which breaks the time reversibility of the equation, and scales with $\rho_f U_0 / H_0$,

corresponding to the flow-induced damping. Under certain conditions, the Q_1 matrix may introduce negative damping and be destabilizing, as discussed in Sec. III D. The last term, scaling with ρ_f / H_0, represents the added-mass effect, and is always present even for zero flow velocity.

C. Ritz method

In this study, the Ritz method was used to solve Eq. (9). The displacement field of the vocal fold structure was approximated as

$$\xi(x,z,t) = \sum_{i=1}^{I} \sum_{k=0}^{K} A_{ik}(t) x^i z^k; \quad \eta(x,z,t) = \sum_{i=1}^{I} \sum_{k=0}^{K} B_{ik}(t) x^i z^k. \tag{19}$$

The generalized coordinates were therefore defined as

$$[q_r] = [A_{10}, A_{11}, \dots, A_{IK}, B_{10}, \dots, B_{IK}]_{N=2 \times I \times (K+1)}. \tag{20}$$

In terms of the generalized coordinates, the fluctuating glottal width was

$$h = -2\xi \bigg|_{\ell_{\text{FSI}}} = -2 \sum_{i=1}^{I} \sum_{k=0}^{K} A_{ik}(t) x^i z^k \bigg|_{\ell_{\text{FSI}}}$$
$$= -2 \sum_{i=1}^{I} \sum_{k=0}^{K} A_{ik}(t) f(z)^i z^k, \tag{21}$$

where $x = f(z)$ describes the prephonatory fluid-structure interface up to the point of flow separation.

D. Linear stability analysis

Substitution of Eq. (18) into Eq. (9) yields

$$(M - Q_2)\ddot{q} + (C - Q_1)\dot{q} + (K - Q_0)q = 0. \tag{22}$$

The equation can be written as

$$\begin{bmatrix} \dot{q} \\ \ddot{q} \end{bmatrix} = \begin{bmatrix} 0 & I \\ -(M - Q_2)^{-1}(K - Q_0) & -(M - Q_2)^{-1}(C - Q_1) \end{bmatrix}$$
$$\times \begin{bmatrix} q \\ \dot{q} \end{bmatrix}, \tag{23}$$

where I is the identity matrix.

Assuming $q = q_0 e^{st}$, the equation is solved as an eigenvalue problem for the eigenvalues s and eigenmodes, q_0. Each eigenvalue corresponds to an eigenmode of the coupled fluid-structure system. To differentiate these eigenmodes from the *in vacuo* modes of the vocal fold structure, they will be referred to as FSI eigenmodes throughout the rest of the paper. The real and imaginary parts of the eigenvalue correspond to the growth rate and the oscillation frequency of the eigenmode, respectively. A positive growth rate indicates that the amplitude will increase with time, indicative of a linearly unstable eigenmode. The mean state of the coupled system is linearly stable if all eigenvalues have negative growth rates. Thus, phonation onset for a particular vocal fold configuration can be studied by solving the eigenvalue problem over a range of subglottal pressures. The phonation threshold pressure would be the subglottal pressure at which the real part

of one eigenvalue first becomes positive. The vibration pattern of the vocal folds at onset would then be given by the corresponding eigenvector.

Note that the mean state (U_0, P_0, H_0) of the coupled system, around which the system equations are linearized, is a function of subglottal pressure. Therefore, for a given subglottal pressure, a nonlinear steady-state problem has to be solved to determine the mean equilibrium state of the coupled system before conducting linear stability analysis around this mean state. For simplicity, the step was skipped in this study and the linear stability analysis was performed on a given mean deformed vocal fold geometry. The mean equilibrium state of the vocal fold was assumed to remain unchanged as the subglottal pressure was increased. With this simplification, the analysis in this study focused on the linear stability of a given pre-phonatory vocal fold geometry.

Note that our model formulation is based on linear theory and therefore does not apply beyond onset. To study the behavior of the system beyond onset, the original nonlinear equations would need to be solved (see, e.g., Holmes, 1977; Guckenheimer and Holmes, 1983).

E. FSI mode decomposition using PCA

Generally the eigenvectors of Eq. (23) are complex. Together with the time dependent factor, e^{st}, a complex eigenvector describes a wave motion along the vocal fold surface, i.e., the vibration of the vocal fold along the FSI interface has no fixed nodal lines. This wave motion can be decomposed into two standing wave components. For the convenience of physical interpretation and direct comparison between different cases, it is desirable to find a consistent decomposition method that can decompose each of the FSI eigenmodes into two spatially and temporally orthogonal components. Using principal components analysis (PCA), each FSI eigenmode was decomposed into two spatially and temporally orthogonal components: a FSI-1 mode and a FSI-2 mode. By definition, the FSI-1 mode captured the largest percentage of the total energy of the original FSI eigenmode. PCA gives an optimal and consistent decomposition: the strongest mode (the FSI-1 mode) obtained using PCA captures more energy than that of any other decomposition (Holmes *et al.*, 1996). The use of PCA would also make it possible to compare the results of this study to previous experimental work in which empirical eigenfunctions of the vocal fold medial surface dynamics were calculated.

For a complex eigenvector q, a matrix could be formed by combining the real and imaginary parts of the eigenvector q

$$X = [\text{Re}(q), \text{Im}(q)] \tag{24}$$

from which a covariance matrix was calculated

$$Cov(X) = X^T M X, \tag{25}$$

where the superscript T denotes the transpose of the matrix. Eigenvalue analysis of the covariance matrix yields two PCA eigenvalues (λ_1, λ_2) and eigenvectors (Ψ_1, Ψ_2). The two eigenvectors provide an orthonormal basis consisting of two vectors whose relative energy weights are given by the rela-
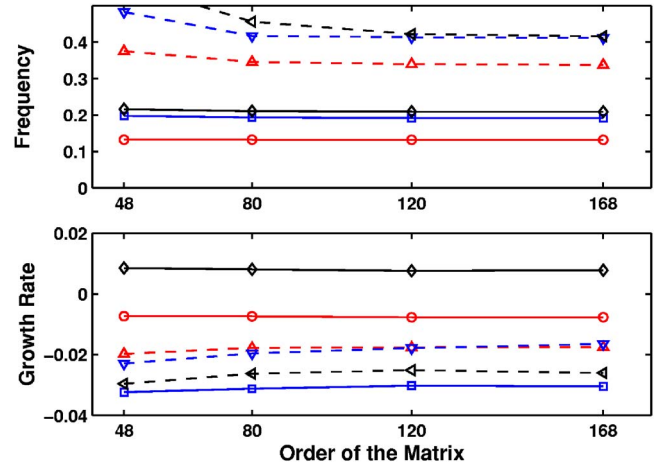
FIG. 2. (Color online) The frequencies and growth rates of the first six eigenvalues (○: first; □: second; ◇: third; △: fourth; ▽: fifth; ◁: sixth) of Eq. (23) as a function of the matrix orders of Eq. (23). The values of model parameters specified in Eq. (28) were used, with a loss factor of 0.1. The order of the matrices in Eq. (23) equals $4 \times I \times (K+1)$. $I=K=3$, 4, 5, and 6 for the four sets of data shown.

tive values of the corresponding eigenvalues. The original eigenvector q can then be projected onto the two orthonormal vectors, giving the final form of the FSI-1 and FSI-2 modes

$$q_{\text{FSI}-1} = \langle q, \Psi_1 \rangle \Psi_1, \quad q_{\text{FSI}-2} = \langle q, \Psi_2 \rangle \Psi_2, \tag{26}$$

where the angular bracket denotes a vector dot product, which is defined as

$$\langle v_1, v_2 \rangle = v_1^{*T} M v_2, \tag{27}$$

where the superscript * denotes complex conjugate, the superscript T denotes transpose, and M is the mass matrix of the vocal fold structure. With this definition, the orthogonality of two vectors is based on the entire vocal fold volume rather than just the medial surface of the vocal folds as implemented in previous laboratory experiments (Berry *et al.*, 2001, 2006).

F. Numerical convergence

Figure 2 shows the eigenvalues of Eq. (23) as a function of the order of the matrices in Eq. (23). Values of the model parameters in Eq. (28) were used, together with a loss factor of 0.1 and a mean jet velocity of 6. For given values of (I, K) for the orders of the polynomials used in Eq. (19), the order of matrices in Eq. (23) equals $4I(K+1)$. It can be seen that convergence was reached for the first four lowest eigenvalues when the order of the matrices was 80 (or above), which corresponded to $I=K=4$ in Eq. (19). In this study, the results presented below were all obtained for this setting. Note that, at this setting, convergence was not reached for higher-frequency eigenvalues. Because higher-frequency eigenmodes tend to dampen out quickly in most real life applications, we focused exclusively on the first four eigenmodes in this study.

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Zhang *et al.*: Physical mechanisms of phonation onset   2283
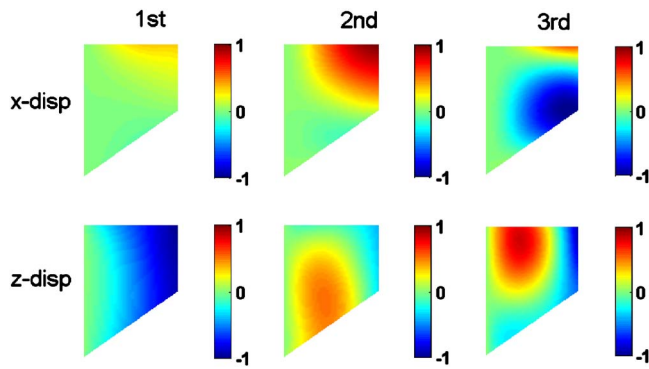
FIG. 3. (Color online) The first three *in vacuo* eigenmodes of the vocal fold, for model parameters specified in Eq. (28). The first and second rows show the *x* component (medial-lateral direction) and *z* component (inferior-superior direction) of the eigenmodes, respectively.

III. RESULTS

This section is organized as follows. Results will first be presented for the case when all three flow terms are excluded from the system Eq. (22) (Sec. III A), which is equivalent to the *in vacuo* condition which was studied in previous investigations (Berry and Titze, 1996; Cook and Mongeau, 2007). Next, the flow-induced stiffness (Sec. III B), inertial (Sec. III C), and damping (Sec. III D) terms will be added sequentially to the equation. Using this procedure, the effects of these three individual terms will be investigated systematically. Through an analysis of the energy flow and the evolution of the eigenvalues, we will show that the flow-induced stiffness was the primary aerodynamic mechanism of phonation onset, involving the excitation and synchronization of two structural eigenmodes. Only under conditions of negligible structural damping and a restricted set of vocal fold geometries did flow-induced damping become the primary mechanism of phonation onset. At moderate and high structural damping, flow-induced stiffness remained the primary aerodynamic mechanism of phonation onset.

For the results presented below, unless otherwise stated, the following nondimensional values for the model parameters were used (see definition in Fig. 1):

$$D = 1.4, \ g = 0.0714, \ T_{\text{inf}} = 1, \ \rho_f = 0.0012, \ \nu = 0.47, \quad (28)$$

which corresponded to the following physical values (the first four are the scaling variables):

$$T = 7 \text{ mm}, \quad \rho_{vf} = 1030 \text{ kg/m}^3, \quad E = 3 \text{ kPa},$$

$$U_{\text{scaling}} = 1.7 \text{ m/s}, \ D = 10 \text{ mm}, \quad g = 0.5 \text{ mm}, \quad (29)$$

$$T_{\text{inf}} = 7 \text{ mm}, \quad \rho_f = 1.2 \text{ kg/m}^3, \quad \nu = 0.47.$$

A. *In vacuo* eigenmodes of the vocal fold

The *in vacuo* eigenmodes of the vocal fold can be obtained by solving Eq. (6) with the right hand side of the equation set to zero. Since no first-order term in time exists in Eq. (6) and both the mass and stiffness matrices are positive definite, the eigenspectrum in this case lies on the imaginary axis, i.e., the growth rates of all eigenvalues are zero.

Figure 3 shows the first three *in vacuo* eigenmodes of the vocal fold. The corresponding eigenfrequencies were 0.102, 0.214, and 0.239 (or 25, 52.2, and 58.2 Hz in physical units for model parameters given in Eq. (29)). The first *in vacuo* eigenmode captured a dominant in-phase vertical motion along the vocal fold medial surface with a weak lateral motion near the superior edge. The second eigenmode captured a dominant in-phase lateral motion along the medial surface. Note that the maximum lateral displacement of the vocal fold along the medial surface occurred at the superior edge. The third eigenmode, in contrast to the first two modes, captured an out-of-phase lateral motion along the vocal fold medial surface. Note that, for the third eigenmode, a nodal line of the lateral displacement existed near the superior edge of the vocal fold. These major features of the first three *in vacuo* eigenmodes of the vocal fold structure are consistent with previous studies (Titze and Strong, 1975; Berry and Titze, 1996).

B. Mechanisms of phonation onset: Flow-induced stiffness

Next we added the flow terms, the matrix Q (Eq. (18)), to the system. For the case of phonation, in which airflow interacts with the vocal fold structure, the density ratio between the flow and structure is very small ($\rho_f \sim 10^{-3}$, Eq. (28)). Therefore, for not so large jet velocities, the flow pressure term Q is generally small compared to the structural mass and stiffness terms in Eq. (22). In particular, the last two terms (Q_1 and Q_2) in Eq. (18) can be neglected and only the flow stiffness term, Q_0, is retained. In the following, we show first the results obtained when only the flow stiffness term was included. This condition is analogous to the linear stability analysis of Ishizaka and colleagues for the two-mass model in which only the flow-induced stiffness was considered (e.g., see pp. 52 and 53 of Ishizaka and Matsudaira, 1972; pp. 233, 238 of Ishizaka, 1981). The other terms in Eq. (16) will be included in subsequent sections of this paper.

1. Eigenmode synchronization

With only the Q_0 term included and structural damping neglected, Eq. (22) becomes

$$M\ddot{q} + (K - Q_0)q = 0. \quad (30)$$

Equation (30) represents a competition between two mechanisms. One is the stabilizing structural restorative forces, the structural stiffness K, and the other is the destabilizing mechanism due to the flow-induced stiffness, Q_0. The solution of Eq. (30) depends on the structure of the flow-induced matrix Q_0. For a symmetric Q_0 matrix, the system loses stability to a static divergence instability (zero frequency instability). This occurs when the flow stiffness term exceeds the structural stiffness term. However, for an asymmetric Q_0 matrix (e.g., due to a nonconservative system), the asymmetric nondiagonal elements in matrix Q_0 induce a cross-mode coupling effect, causing two neighboring eigenvalues to approximate each other and eventually collide. When this coupling effect is strong, the system loses stability to a coupled-mode flutter instability (Weaver, 1974; Holmes, 1977; Auregan and
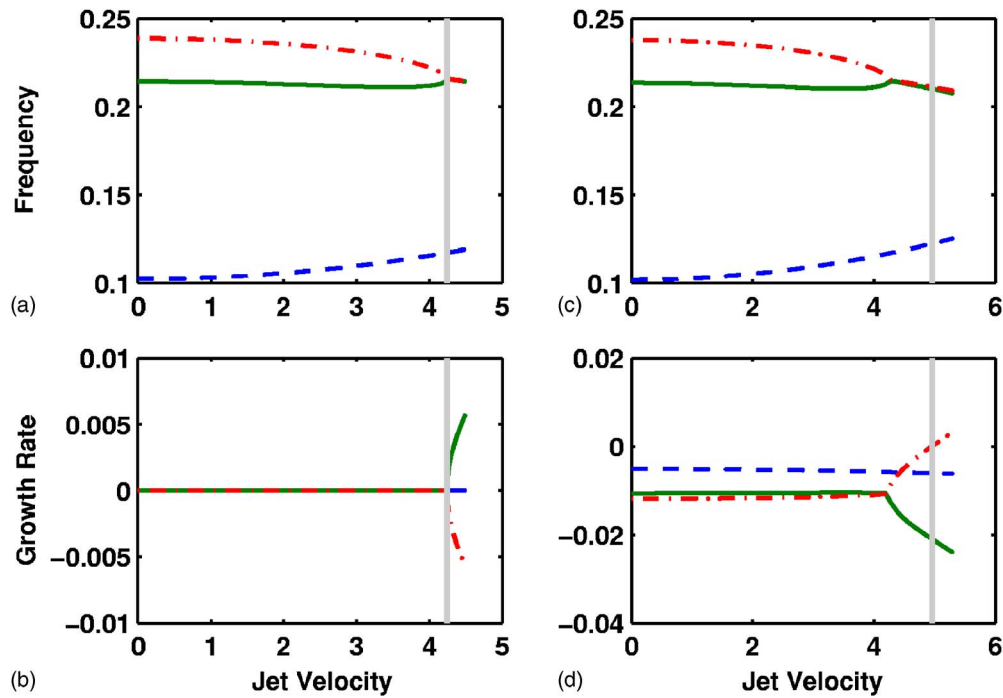
FIG. 4. (Color online) The frequencies and growth rates of the first three eigenvalues (– –: first; ——: second; -.-.-: third) of the coupled fluid-structure system as a function of the jet velocity. The vertical line indicates the point of onset. Model parameters specified in Eq. (28) were used. Only the flow stiffness term Q_0 of the three flow-induced terms was included in Eq. (22). Figures 4(a) and 4(b): $\sigma=0$, the corresponding eigenvalue movement in the complex half plane is shown in Fig. 5(a); Figs. 4(c) and 4(d): $\sigma=0.1$, the corresponding eigenvalue movement in the complex half plane is shown in Fig. 5(b).

Depollier, 1995; Cullen *et al.*, 2000), as further illustrated below.

Figures 4(a) and 4(b) show the first three eigenfrequencies of Eq. (30) and their associated growth rates as a function of jet velocity. At zero velocity, Eq. (30) reduces to Eq. (6) with the eigenvalues corresponding to the *in vacuo* eigenvalues of the vocal fold. As the jet velocity increased, the second and third eigenvalues approached each other in frequency while the growth rates for all remained zero. At a threshold value of the jet velocity (phonation onset), the second and third eigenfrequencies merged into one frequency, while at the same time one growth rate became positive and the other negative. Figure 5(a) shows the movement of the second and third eigenvalues in the complex half plane. Note that, although eigenvalues always come in complex conjugate pairs, only the eigenvalues with positive imaginary part are shown in the complex half plane in Fig. 5. Before onset, all the eigenvalues lay on the imaginary axis. As the flow velocity increased, the second and third eigenvalues, mainly the third eigenvalue, moved towards each other due to the coupling effect of the flow. At onset, the two eigenvalues collided and split into two eigenvalues with the same frequency but with nonzero growth rates of opposite sign. Therefore, in this case, onset occurred via a 1:1 resonance mechanism (Guckenheimer and Holmes, 1983) of the second and third eigenvalues.

Note that the asymmetric nature of the matrix Q_0 in our model was mainly due to the imposed flow separation (see Eqs. (8) and (17)). For a potential flow formulation without flow separation, the matrix Q_0 would be nearly symmetric, and Eq. (30) would lose stability first to a static divergence of the first eigenmode, rather than to a coupled-mode flutter

instability. In fact, coupled-mode flutter is facilitated by any factor or mechanism which makes the matrix Q_0 asymmetric. For example, our simulations showed that coupled-mode flutter also occurred first for Eq. (30) for sufficiently large viscous flow resistance but without flow separation.

The structural restorative force in the vocal folds scales with $\omega_0^2 \xi$, where ω_0 is the *in vacuo* eigenfrequency. The flow-induced stiffness term, and therefore the coupling strength, roughly scales with $\rho_f U_j^2 \xi / g$. Therefore, a balance between the structural stiffness and the flow-induced stiffness yields the following expressions or proportionalities for the jet velocity and subglottal pressure at onset:

$$U_j \sim \omega_0 \sqrt{\frac{g}{\rho_f}}; \quad P_{0,in} \sim \omega_0^2 g. \tag{31}$$

The *in vacuo* eigenfrequency ω_0 is a function of the material properties of the vocal folds, vocal fold geometry, and boundary conditions (Berry and Titze, 1996; Cook and Mongeau, 2007).

2. Vocal fold vibration at onset

The eigenmode with the largest growth rate, or the critical eigenmode, was analyzed to investigate vocal fold vibration characteristics at onset. The phonation frequency, or the imaginary part of the eigenvalue of the critical eigenmode, was in between the second and third *in vacuo* eigenfrequencies, slightly closer to the second *in vacuo* eigenfrequency, as shown in Fig. 4(a). The corresponding complex eigenvector was decomposed into two spatially and temporally orthogonal FSI modes, as described in Sec. II E. The vibration pattern of the vocal fold for each FSI mode can be obtained by
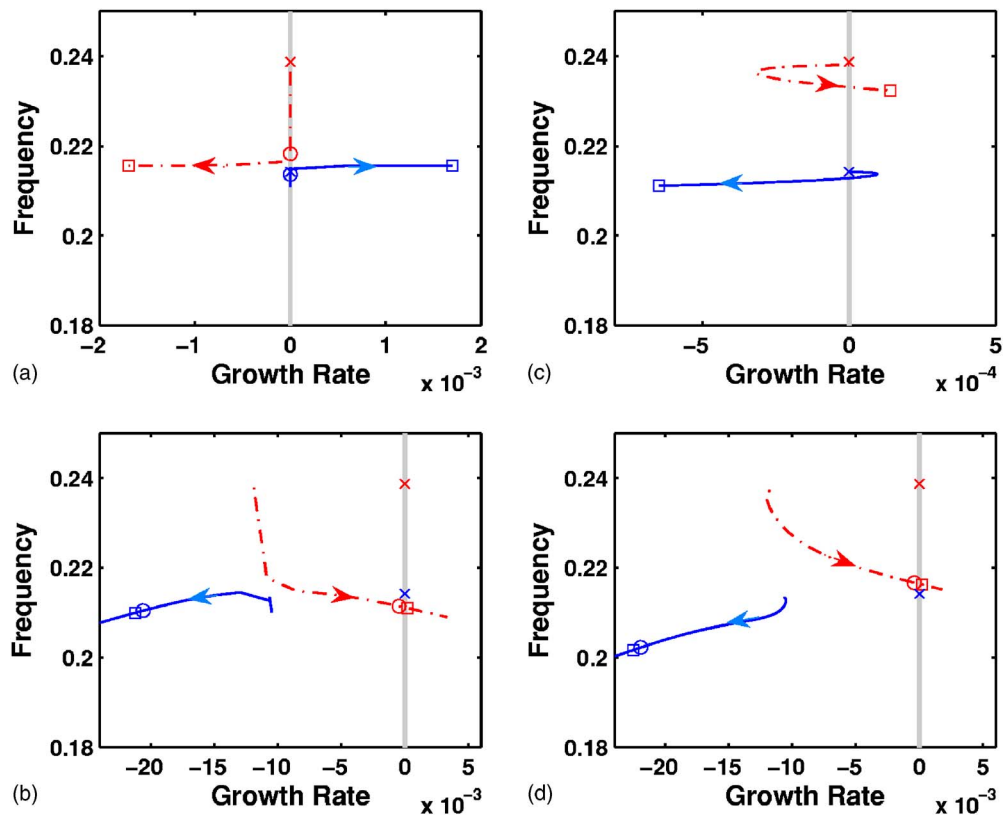
FIG. 5. (Color online) Movement of the second and third eigenvalues in half complex plane. Model parameters specified in Eq. (28) were used. (a) with only the flow stiffness term Q_0 of the three flow-induced terms included, $\sigma=0$; (b) with only the flow stiffness term Q_0 of the three flow-induced terms included, $\sigma=0.1$; (c) with all three flow-induced terms included, $\sigma=0$; (d) with all three flow-induced terms included, $\sigma=0.1$; ×: *in vacuo* eigenvalues; ○: eigenvalues before onset; □: eigenvalues after onset. The arrows indicate the direction of the movement of the eigenvalues as the jet velocity increases. The vertical lines indicate the imaginary axis.

substituting the corresponding eigenvector into Eq. (19), and is shown in Fig. 6.

The dominant FSI mode, FSI-1, captured an out-of-phase lateral motion and an equally strong in-phase vertical motion along the medial surface. The second FSI mode, FSI-2, captured a dominant in-phase lateral motion and a relatively weak vertical motion along the medial surface. Neither of the FSI modes closely resembled the *in vacuo*
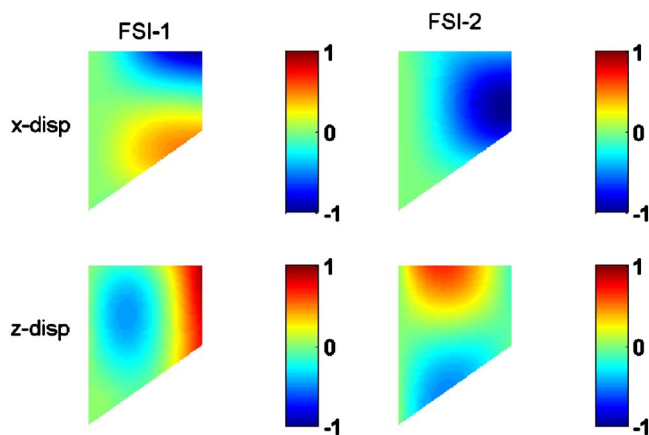
FIG. 6. (Color online) Vibration patterns of the FSI-1 (left) and FSI-2 (right) modes at onset for case shown in Fig. 5(a) ($Q=Q_0 q$, model parameters given by Eq. (28), $\sigma=0$). The first and second rows show the x component (medial-lateral direction) and z component (inferior-superior direction) of the eigenmodes, respectively.

eigenmodes of the vocal fold. In fact, the FSI-1 and FSI-2 were to some degree similar to the third and second *in vacuo* eigenmodes with the vibration pattern (within the entire vocal fold volume) shifted inferiorly, respectively. For example, in the FSI-1 mode, a horizontal (medial-lateral) nodal line of the lateral displacement (x component) of the vocal fold existed around the center of the superior section ($z>0$) of vocal fold in the inferior-superior direction, as compared to the nodal line near the superior edge for the third *in vacuo* eigenmode of the vocal fold (Fig. 3). In the FSI-2 mode, the maximum lateral displacement along the medial surface occurred around the center of the superior section of the vocal fold surface, shifted inferiorly from its location at the superior edge for the second *in vacuo* eigenmode. To quantify the similarity between the FSI modes and the *in vacuo* modes, correlations (vector dot product) between the two FSI modes and the *in vacuo* eigenmodes were calculated based on Eq. (27). The correlations between the FSI-1 mode and the first three *in vacuo* eigenmodes were 31.3%, 69.1%, and 65.1%, respectively. For the FSI-2 mode, its correlations with the first three *in vacuo* eigenmodes were 13.5%, 64.75%, and 75.0%, respectively. The correlation between the two FSI modes and higher-order *in vacuo* eigenmodes was nearly zero so that both FSI modes could be reproduced by a combination of the first three *in vacuo* eigenmodes with an accuracy above 99%.
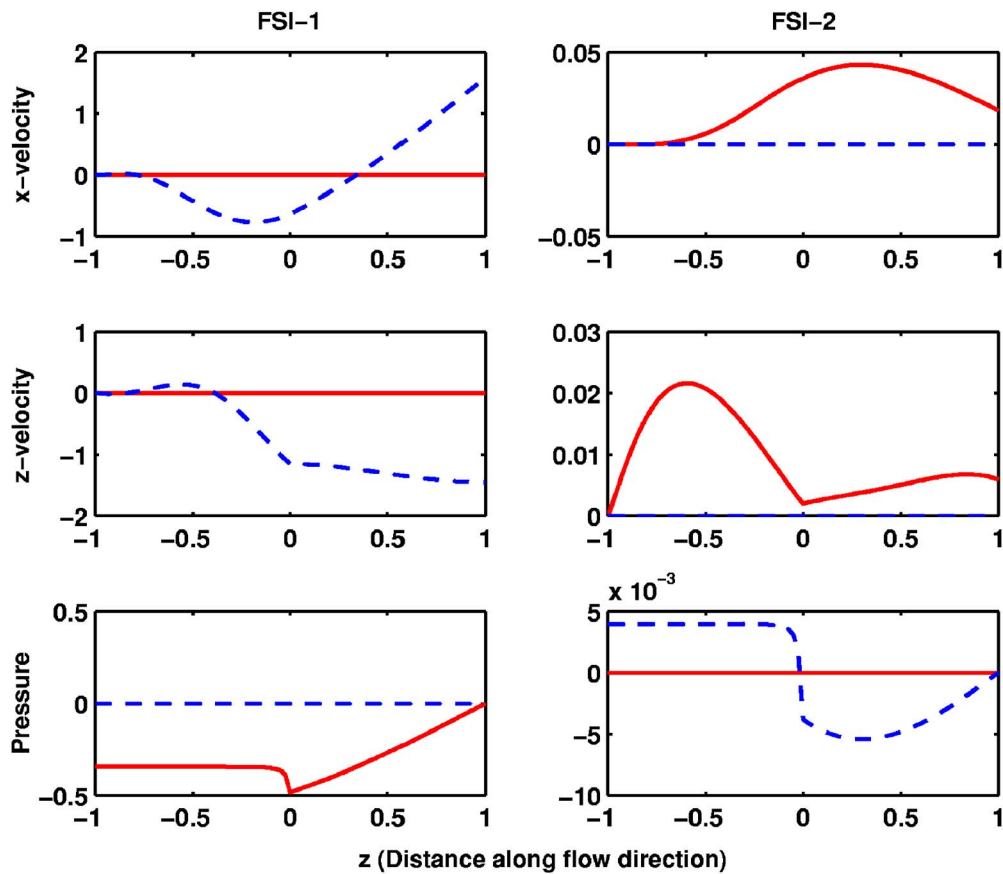
FIG. 7. (Color online) The vocal fold velocity in medial-lateral and inferior-superior directions and the flow pressure along the vocal fold surface corresponding to the FSI-1 and FSI-2 modes as shown in Fig. (6). ——: real part; – –: imaginary part.

At onset, the relative energy weights (λ_1, λ_2) of the FSI-1 and FSI-2 modes were 99.9% and 0.1%, respectively, which indicates that the vocal fold vibration was essentially a single-mode, standing wave motion rather than a traveling wave motion. Ideally, at the exact point of onset, all the eigenvalues of Eq. (30) are still on the imaginary axis, including the colliding pair. The eigenvectors corresponding to the colliding pair of eigenvalues would be real. In this case (without structural damping), the space spanned by the two FSI modes is reduced to a one-dimensional space, i.e., a single mode rather than a traveling wave motion. The relative energy weights of the two FSI modes would be 100% and 0%, representing a single-mode vibration pattern. Our numerical analysis did not catch the exact onset point so that the relative weights deviated from the exact theoretical values. When structural damping is included, as shown later in Sec. III B 4, the onset would be delayed until after the eigenvalue collision. The eigenvector of the critical eigenmode would therefore be complex, describing a traveling wave as generated by the superposition of two FSI modes. In the complex full plane, this eigenmode corresponded to a complex conjugate pair of eigenvalues, with a complex conjugate pair of eigenvectors. The space spanned by the two FSI modes is therefore two dimensional. The weight of the FSI-1 mode would decrease from 100% and two FSI modes would then be needed to describe the vocal fold vibration. With

increasing structural damping, the two eigenvectors will further deviate away from each other, leading to more comparable weights for the two FSI modes (Sec. III B 4).

3. Energy transfer

Figure 7 shows the vocal fold velocity and pressure distribution along the medial surface of the vocal fold for the two FSI modes slightly above onset. For both modes, the pressure was 90° out of phase with the velocity (one was nearly pure real and the other was nearly pure imaginary) except for a very small in-phase component. The first FSI mode, FSI-1, captured an out-of-phase lateral motion along the medial surface, creating an alternating convergent/divergent glottis near the superior part of the glottal channel $(0 < z < 1)$. Due to the zero pressure boundary condition at the superior edge of the glottal channel, this motion created an alternating in-phase increase/decrease in the pressure along the glottal channel. On the other hand, the FSI-2 mode induced an out-of-phase pressure distribution along the glottal channel.

From the distribution of the vocal fold velocity along the medial surface and the pressure along the glottal channel, the energy flow from the fluid into the structure was calculated as

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Zhang *et al.*: Physical mechanisms of phonation onset   2287
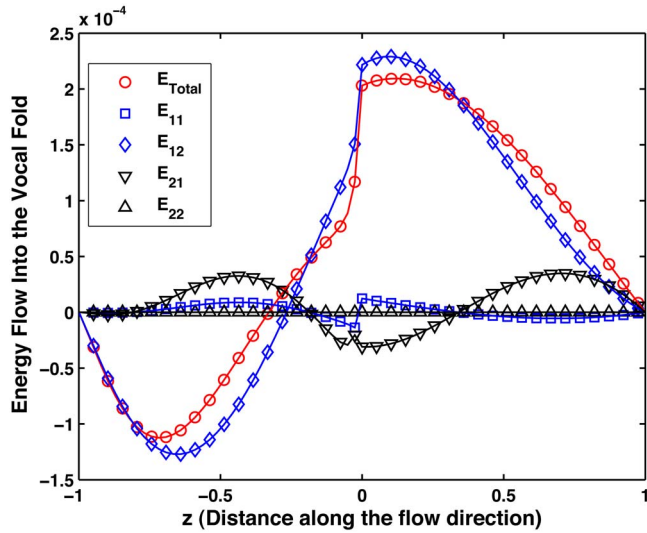
FIG. 8. (Color online) Distribution of total energy flow into the vocal fold along the vocal fold surface and its decomposition into contributions of two self-mode and two cross-mode interaction terms for the case shown in Fig. 5(a) (model parameters given in Eq. (28), $Q=Q_0q$, and $\sigma=0$, slightly above onset). See Eqs. (32) and (33) for the definition of different energy terms.

$$E_{total} = -\frac{1}{2}\,\mathrm{Re}\left(p\dot{\xi}^* n_x + p\dot{\eta}^* n_z\right)\Big|_{\ell_{FSI}} = E_{11} + E_{12} + E_{21} + E_{22},$$

(32)

where

$$E_{11} = -\frac{1}{2}\,\mathrm{Re}\left(p_{fsi-1}\dot{\xi}^*_{fsi-1} n_x + p_{fsi-1}\dot{\eta}^*_{fsi-1} n_z\right)\Big|_{\ell_{FSI}},$$

$$E_{12} = -\frac{1}{2}\,\mathrm{Re}\left(p_{fsi-1}\dot{\xi}^*_{fsi-2} n_x + p_{fsi-1}\dot{\eta}^*_{fsi-2} n_z\right)\Big|_{\ell_{FSI}},$$

(33)

$$E_{21} = -\frac{1}{2}\,\mathrm{Re}\left(p_{fsi-2}\dot{\xi}^*_{fsi-1} n_x + p_{fsi-2}\dot{\eta}^*_{fsi-1} n_z\right)\Big|_{\ell_{FSI}},$$

$$E_{22} = -\frac{1}{2}\,\mathrm{Re}\left(p_{fsi-2}\dot{\xi}^*_{fsi-2} n_x + p_{fsi-2}\dot{\eta}^*_{fsi-2} n_z\right)\Big|_{\ell_{FSI}}.$$

In Eq. (32), the total energy flow is decomposed into two self-mode interaction terms and two cross-mode interaction terms. Figure 8 shows the distribution of the total energy flow along the glottal channel. In this case, energy flow was negative along the inferior part of the glottal channel ($z < 0$), indicating that energy was transferred from the structure to the fluid. In the superior part of the glottal channel ($z > 0$), energy flow was reversed and energy was transferred from the fluid into the structure at a slightly greater rate than that at which energy was extracted from the inferior vocal fold structure. In total a net positive energy transfer was established from the fluid into the vocal fold structure. Figure 8 also shows the individual contributions of the self-mode and cross-mode interaction between the two FSI modes. The contributions of the two self-mode interaction terms were small, as expected for a pressure field nearly 90° out of phase with the velocity. Of the two cross-mode terms,

E_{12} was the most dominant term, which represented the interaction between the pressure field induced by the out-of-phase FSI-1 mode and the velocity field induced by the in-phase FSI-2 mode. Integrated along the glottal channel, the four energy terms ($E_{11}, E_{12}, E_{21}, E_{22}$) contributed 1.14%, 83.71%, 15.14%, and 0.01% to the total energy flow, respectively.

The above analysis further clarified the onset mechanism from the point of view of energy transfer. As discussed in Sec. III B 2, a major difference between the eigenvalues before and after onset is that the complex eigenvalues after onset resulted in complex eigenvectors, describing a wave motion or the synchronization of two FSI modes at precisely the same frequency. Without this synchronization, the energy flow would be zero as the pressure induced by the same mode would be always 90° out of phase with the vocal fold velocity. It is only because of the synchronization of two modes that the total pressure field had an in-phase component with the total vocal fold velocity field (Fig. 7), no matter how small the FSI-2 mode was in amplitude compared to the FSI-1 mode. In particular, the most effective energy transfer contribution came from the interaction between the pressure field of the FSI-1 mode (capturing an out-of-phase lateral motion) with the velocity field of the FSI-2 mode (capturing an in-phase lateral motion). This is consistent with previous understanding (Ishizaka, 1981, 1988; Titze, 1988; Berry and Titze, 1996) that two eigenmodes of the vocal fold vibration "entrained" to create favorable aerodynamic conditions for phonation: one eigenmode shapes the top of the glottal airway to be alternatingly convergent/divergent during one glottal oscillation cycle, therefore modulating the intraglottal pressure, while the other eigenmode governs the net lateral tissue velocity and modulates the flow velocity. The two eigenmodes interacted to facilitate energy transfer from the airflow to the vocal fold tissue, as shown in Fig. 8.

Since no damping terms (either structural or flow induced) were included in the present example, the energy flow at onset must be identically zero rather than nonzero, as in the above energy analysis. However, since the eigenvalue problem (Eq. (30)) was solved numerically in this study, the exact point of onset was not captured. Therefore, the onset jet velocity as identified in this study was always slightly larger. While our model formulation no longer applies after onset, it will be shown below that the general conclusions (the vibration patterns of the two FSI modes and energy transfer) made in this subsection still hold for cases with nonzero structural damping, in which a nonzero energy flow was present to balance the structural dissipation.

4. Effects of structural damping

Inclusion of structural damping shifted the eigenspectrum of the system to the left in the complex half plane, as shown in Fig. 5(b) for the same set of model parameters but with a loss factor of $\sigma=0.1$. Figures 4(c) and 4(d) show the corresponding frequency and growth rate of the first three eigenmodes. The physical mechanism of onset remained the same: two eigenvalues approached each other and then separated again, one of which eventually crossed the imaginary axis as the jet velocity continued to increase, causing the
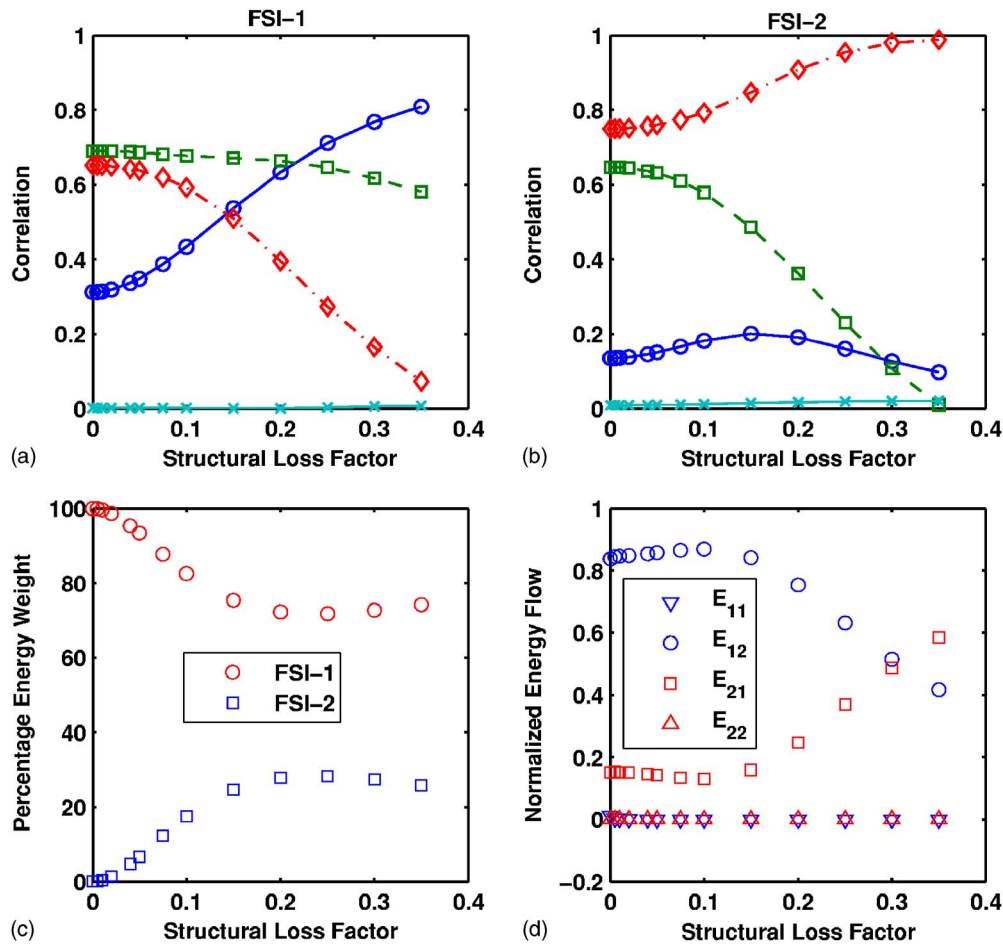
FIG. 9. (Color online) Correlation between the FSI-1 mode (a) and FSI-2 mode (b) and the first four *in vacuo* eigenmodes (○: first; □: second; ◇: third; ×: fourth *in vacuo* eigenmodes), relative energy weights of the two FSI modes (c) and normalized contributions to the total energy flow by individual interaction terms (d) as a function of the structural loss factor. Model parameters were given in Eq. (28), $Q = Q_0 q$.

corresponding eigenmode to become linearly unstable. However, in contrast to the case of Fig. 5(a), the two eigenvalues in Fig. 5(b) approached but never actually collided with each other. Note the similarity between Figs. 4(c) and 4(d) in this paper and Fig. 5 of (Ishizaka, 1981, p. 238). Using two different computational models, both figures have apparently captured the same physical mechanism.

Figure 9 shows the correlations between the two FSI modes and the first four *in vacuo* eigenmodes as a function of the loss factor. Both FSI modes were a mixture of mainly the first three *in vacuo* eigenmodes of the vocal folds, with negligible correlations with the fourth and higher *in vacuo* eigenmodes (Figs. 9(a) and 9(b)). For small and moderate structural damping, the vocal fold vibration patterns captured by the two FSI modes were qualitatively similar to the case without structural damping. The FSI-1 mode captured an out-of-phase lateral motion, with a nodal line along the center of the medial surface, and an in-phase vertical motion along the medial surface of the vocal fold. The FSI-2 mode captured a dominant, in-phase lateral motion with a weak, in-phase vertical motion along the vocal fold medial surface. As the loss factor increased, the correlation with the third *in vacuo* eigenmode slightly decreased (increased) for the FSI-1 (FSI-2) mode. For the FSI-1 mode, this caused the nodal line of the lateral displacement to continuously shift inferiorly.

For the FSI-2 mode, this caused the location of the maximum lateral displacement along the medial surface to shift inferiorly. There was a threshold value of the structural loss factor (around $\sigma = 0.25$) at which the vibration patterns captured by the two FSI modes became qualitatively similar. Beyond this threshold, a switchover occurred in the vibration patterns captured by the two FSI modes. The FSI-1 mode now captured an in-phase lateral motion while the FSI-2 mode captured an out-of-phase lateral motion.

Figure 10 shows the onset jet velocity and onset frequency as a function of the loss factor, for model parameters specified by Eq. (28). The onset frequency in this case slightly decreased with the loss factor. With structural damping included, more energy transfer was needed to overcome the structural dissipation, delaying onset to a higher value of the jet velocity. This additional energy transfer was achieved by the increasing relative weight of the FSI-2 mode with the loss factor (Fig. 9(c)). With more comparable weights of the two FSI modes, the total flow pressure became more in phase with the vocal fold velocity, which increased the energy transfer efficiency and allowed more energy transfer to overcome structural dissipation.

The energy transfer mechanisms remained similar to the case without structural damping. Except for ranges of loss factors in which the two FSI modes became similar, the ma-

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Zhang *et al.*: Physical mechanisms of phonation onset 2289
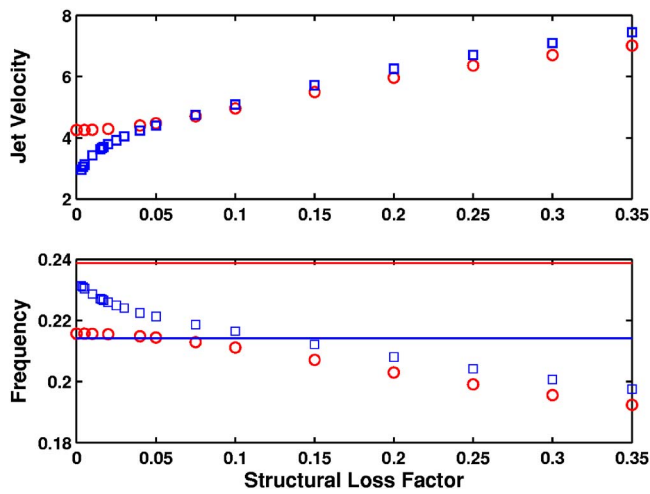
FIG. 10. (Color online) Onset jet velocity (top) and frequency (bottom) as a function of the structural loss factor, for model parameters given in Eq. (28). $\bigcirc$: $Q=Q_0 q$; $\square$: $Q=Q_2\ddot{q}+Q_1\dot{q}+Q_0 q$. The two horizontal lines indicate the second and third *in vacuo* eigenfrequencies.

FIG. 11. (Color online) The frequencies and growth rates of the first three eigenmodes (– –: first; ——: second; -.-.-: third) as a function of the jet velocity for three cases in which all three flow terms were included ($Q = Q_2\ddot{q}+Q_1\dot{q}+Q_0 q$). (a-b): model parameters given by Eq. (28), $\sigma=0$; (c-d): model parameters given by Eq. (28) except that $T_{inf}=1.5$, $\sigma=0$; (e-f): model parameters given by Eq. (28), $\sigma=0.1$. The vertical lines indicate point of onset.

jor contribution of energy flow still came from the interaction between the pressure field induced by the out-of-phase FSI mode and the vocal fold motion of the in-phase FSI mode (Fig. 9(d)). The other cross-mode interaction term also positively contributed to the total energy flow. The contributions from the same-mode interaction were negligible.

C. Effects of flow-induced inertia

As discussed in Sec. II B, the flow-induced inertia term, p_2, is the weakest term of the three parts in Eq. (16), due to the small density ratio between air and the vocal fold. When included, the eigenspectrum remained nearly the same as when it was excluded, with only slight changes in the frequencies. The difference was nearly unnoticeable except for very small glottal half widths, which would change the relative weight of the flow-induced inertia term with respect to the structural mass term (Eq. (17)).

D. Effects of flow-induced damping

1. Destabilization effects

The flow-induced damping term, p_1, is a first-order term in time. It breaks the time reversal symmetry of Eq. (30), and, depending on the structure of the matrix Q_1, may directly destabilize the coupled fluid-structure system. Figures 11(a) and 11(b) show the first three eigenvalues as a function of increasing jet velocity. The model parameters were the same as in Sec. III A. (Eq. (28), and zero structural damping), but with all three flow terms, Q_0, Q_1, and Q_1 included in Eq. (22). The corresponding movement of the second and third eigenvalues in the complex half plane is shown in Fig. 5(c). Unexpectedly we found that the second eigenmode immediately became unstable when a nonzero flow was applied, while at the same time the first and third eigenmodes were stabilized. In other words, the onset jet velocity was infinitesimally small.

This seemingly unlikely result was due to the presence of an irreversible perturbation (the flow-induced damping term) in an otherwise reversible system. The irreversible perturbation introduces a first-order term in time whose matrix coefficient, the flow damping term Q_1 in our case, is no longer positive-definite so that destabilization occurs. In this case, a negative damping was introduced in the second eigenmode and a positive damping in the first and third eigenmodes (Fig. 11(b)). Due to this destabilization effect, the onset of the perturbed system would in general be lower than in the unperturbed system. In our case, the onset jet velocity was decreased to an infinitesimally small number. The destabilization of reversible systems due to irreversible perturbations has been extensively investigated (e.g., Kounadis, 1992; O'Reilly *et al.*, 1996; Kirillov, 2005). O'Reilly *et al.* (1996) examined the destabilization of the equilibria of a reversible system when an infinitesimally small first derivative term in time (the flow damping term in our case) was introduced to the system. Based on the matrix invariants of system governing equations, they established the necessary and sufficient conditions for a two degree-of-freedom system to lose linear stability due to the addition of an irreversible perturbation. Using similar criteria, the eigenvalue movement as shown in Figs. 11(a) and 11(b) can be predicted, i.e., one eigenmode was destabilized while the other stabilized.

For this example without structural damping (Figs. 11(a), 11(b), and 5(c)), since the second eigenmode was destabilized at an infinitesimally small flow velocity, the eigenmodes interacted minimally with each other so that the critical eigenmode was almost the same as the corresponding *in vacuo* eigenmode of the vocal fold structure. Indeed, a close examination showed that the two FSI modes at a very small jet velocity (which is of course slightly beyond onset) were highly similar to the third and the second *in vacuo* eigenmodes of the vocal fold structure, respectively. In fact, the correlation between the FSI-1 mode and the second *in vacuo* eigenmode was above 99%, and the FSI-2 mode was 84% and 53% correlated with the third and first *in vacuo* eigenmodes, respectively. In addition, the FSI-1 mode captured
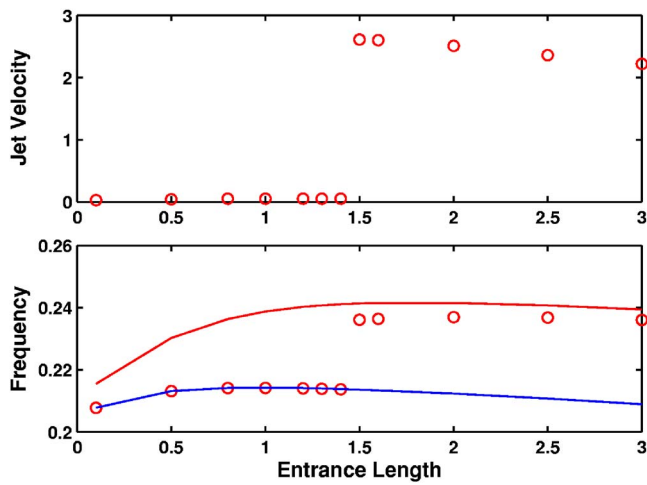
FIG. 12. (Color online) Onset jet velocity (top) and frequency (bottom) as a function of the length of the entrance section T_{inf}, for model parameters given in Eq. (28), $Q=Q_2\ddot{q}+Q_1\dot{q}+Q_0q$ for $\sigma=0$. The two solid lines indicate the second and third *in vacuo* eigenfrequencies.

more than 99% of the total energy. As expected, the critical eigenmode was essentially the second *in vacuo* eigenmode with nearly zero contribution from the other *in vacuo* eigenmodes. The pressure field in the glottal channel and the vocal fold velocity field along the medial surface showed that, for both FSI modes, the pressure fields were in phase with the velocity fields so that the self-mode interaction due to the FSI-1 mode (due to its 99% weight) dominated the total energy flow. In other words, the destabilizing effect of the flow-induced damping is a one-mode instability (the second *in vacuo* eigenmode destabilized by flow-induced damping), as compared to the synchronization of two eigenmodes in Sec. III B.

2. Structure of the Q_1 matrix

The destabilization effect is highly dependent on the structure of the flow damping term Q_1, which is again dependent on many factors such as the vocal fold geometry. For example, Figs. 11(c) and 11(d) show the frequency and growth rate as a function of the jet velocity for a case with the same conditions as the case in Figs. 11(a) and 11(b), except the length of the entrance section ($z<0$) of the glottal channel T_{inf} was increased from 1 to 1.5. The first three eigenmodes were all stabilized when the flow was applied, and the system was now linearly stable. Figure 12 shows the onset jet velocity and frequency as a function of the length of the entrance section T_{inf}. A threshold value of T_{inf} (about 1.4) existed above which the onset jet velocity jumped to a finite number. Beyond this threshold, the flow-induced damping became stabilizing so that the system was no longer linearly unstable at very small jet velocities (Fig. 11(d)). Correspondingly, the critical eigenvalue changed from the second to the third eigenvalue, as shown in Fig. 12. Note that the onset jet velocity for values of T_{inf} above the threshold was still much lower than it would have been had the flow-induced damping term been excluded, as discussed further below.

3. Effects of structural damping

The destabilizing effect of the flow-induced damping was suppressed when structural damping was included to the coupled system. Figures 11(e) and 11(f) show the frequency and growth rate as a function of the jet velocity for the same case as in Figs. 11(a) and 11(b) but with a loss factor of 0.1. The corresponding movement of the second and third eigenvalues in the complex half plane is shown in Fig. 5(d). When a nonzero flow was applied, the second and third eigenvalues started in opposite directions, under the influence of the flow-induced damping. As the jet velocity increased, the two eigenvalues eventually reversed directions under the influence of the flow stiffness term. The system remained stable until the jet velocity reached a finite value. As the jet velocity further increased, the flow stiffness term dominated and eventually destabilized the coupled system (onset). Although a relatively large loss factor (0.1) was used in this example, a much smaller loss factor would also suppress the destabilizing effect of the flow-induced damping, as discussed later. For moderate and high structural damping, the effects of the flow-induced damping on phonation onset became secondary, and the onset was again primarily determined by the properties of the flow-induced stiffness matrix, as discussed in Sec. III B.

However, similar to that of the structural damping (Sec. III B 4) but to a much larger extent, the presence of the flow damping term destroyed the interaction between the second and third eigenvalues, which would have collided on the imaginary axis and led to onset as in Sec. III B. Phonation onset occurred well before the eigenvalues sufficiently approximated each other (compare with Figs. 5(a) and 5(b)). In fact, the two eigenvalues only minimally approached each other in the complex half plane and never collided.

Figure 10 shows the onset jet velocity and onset frequency (denoted by square symbols) as a function of the loss factor for the model parameters specified by Eq. (28), with all three terms of the flow pressure included in Eq. (22). At a very small value of the loss factor, the onset jet velocity jumped from an infinitesimally small one (not shown in Fig. 10) to finite values (about 3). After that, the onset jet velocity increased smoothly with increasing loss factor, as in the case when only flow stiffness was included. Compared to the case with only the flow stiffness term included, two regions of the loss factor can be identified. For loss factors smaller than about 0.06, the onset jet velocity was lowered when the flow damping was added due to the destabilizing effect of the flow damping term. For loss factors larger than 0.06, the onset jet velocity was higher when the flow damping term was included. Obviously, the flow-induced damping changed from destabilizing to stabilizing at large structural damping.

Figure 13 shows the correlations between the two FSI modes and the first four *in vacuo* eigenmodes, the relative energy weights of the two FSI modes, and the energy flow decomposition as a function of the loss factor, for cases in which all three terms of the flow pressure were included in Eq. (22). Similar observations can be made as those in Sec. III B 4. Both FSI modes were a mixture of the first three *in vacuo* eigenmodes of the vocal folds. For small and moderate structural damping, the FSI-1 mode captured an out-of-

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Zhang *et al.*: Physical mechanisms of phonation onset   2291
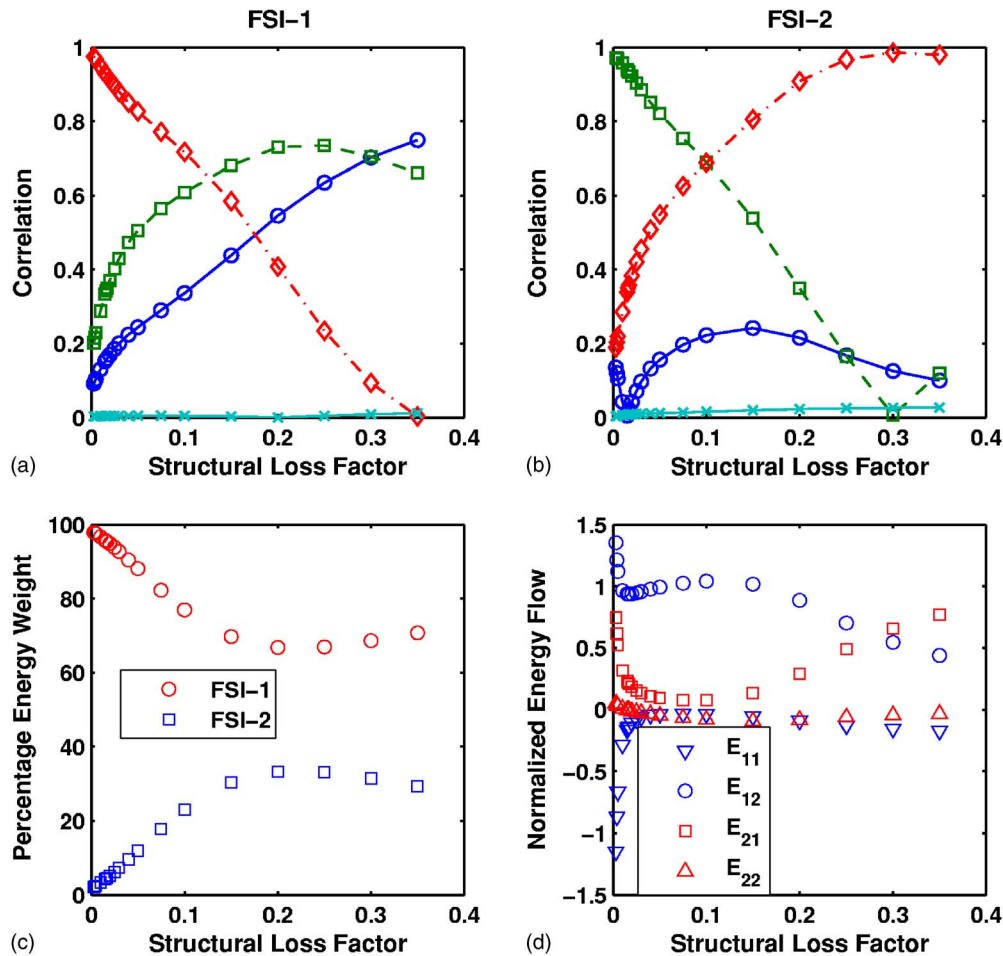
FIG. 13. (Color online) Correlation between the FSI-1 mode (a) and FSI-2 mode (b) and the first four *in vacuo* eigenmodes (○: first; □: second; ◇: third; ×: fourth *in vacuo* eigenmodes), relative energy weights of the two FSI modes (c), and normalized contributions to the total energy flow by individual interaction terms (d) as a function of the structural loss factor. Model parameters were given in Eq. (28), $Q=Q_2\ddot{q}+Q_1\dot{q}+Q_0q$.

phase lateral motion, while the FSI-2 mode captured an in-phase lateral motion. The vibration patterns changed around $\sigma=0.25$ so that, for higher structural damping, the FSI-1 captured an in-phase lateral motion while the FSI-2 mode captured an out-of-phase motion. The value of the loss factor at the switchover was not much different from the case when only the flow stiffness term was included. As the structural damping increased, the relative weight of the FSI-1 mode decreased while that of the FSI-2 mode increased. Note that for moderate to high structural damping, the relative weights of the two FSI modes (about 70% and 30%, respectively) were similar to those of the surface-based empirical eigenfunctions observed in laboratory experiments (Berry *et al.*, 2001; Zhang *et al.*, 2006b). Concerning the energy transfer mechanism, as the structural damping increased, the pattern of energy flow decomposition approached that of the case when only flow stiffness was included. That is, the dominant contribution to the total energy transfer came from the interaction between the pressure field induced by the out-of-phase FSI mode and the vocal fold motion of the in-phase FSI mode.

IV. DISCUSSION

In this study, the primary mechanism of phonation onset was flow-induced stiffness, rather than flow-induced damp-

ing. In particular, the asymmetric nature of the flow-induced stiffness matrix induced two neighboring eigenvalues to approximate each other and eventually merge, initiating phonation onset. In this study, the asymmetry of the flow-induced stiffness matrix was mainly caused by flow separation. However, such an asymmetry may also be induced by other non-conservative mechanisms such as viscous flow resistance. Only under conditions of negligible structural damping and a restricted set of vocal fold geometries did flow-induced damping become the primary mechanism of phonation. At moderate to high structural damping, although phonation onset was not directly related to an eigenmode-synchronization, the primary mechanism of phonation onset remained to be the synchronization effect of the flow-induced stiffness.

Some of the initial results of this investigation closely parallel the linear stability analysis of the two-mass model performed by Ishizaka and colleagues (Ishizaka and Matsudaira, 1972; Ishizaka, 1981, 1988), particularly with respect to the role of the flow-induced stiffness term in facilitating eigenmode synchronization and initiating phonation. These studies underscore the necessity of retaining at least two degrees-of-freedom in lumped-element models of phonation. For a single degree-of-freedom system (e.g., the one-mass model or any degenerate two-mass model), eigenmode synchronization mechanism is excluded *a priori*. For example,

in the one-mass model, the only mechanism of phonation relies on externally introduced damping (e.g., by coupling to an acoustic resonance). Without externally introduced damping, only a static divergence instability (due to the interaction between the single eigenvalue with its complex conjugate) is possible in a single degree-of-freedom system.

It is difficult to compare the results of the present investigation with results of Titze (1988). In part, this is because the present study used a self-oscillating, two-dimensional, fluid-structure interaction system, while Titze (1988) used a two degree-of-freedom system with prescribed phase relationship between the two eigenmodes. Therefore, while Titze (1988) investigated the consequences of eigenmode synchronization, he failed to investigate the mechanisms of eigenmode synchronization. In particular, Titze (1988) showed that the synchronization of two structural eigenmode produced an effective negative damping term in the resulting single degree-of-freedom system. However, this concept of negative damping provides no insight regarding how eigenmode synchronization might be initiated and facilitated by the aeroelastic properties of the coupled fluid-structure-interaction system. Furthermore, the concept of modeling eigenmode synchronization as an effective negative damping, similar to the flow-induced damping introduced by sub- and supra-glottal acoustics (Titze, 1988; Zhang *et al.*, 2006a), may be misleading. Indeed, the present study shows that flow-induced stiffness, rather than flow-induced damping, was the primary mechanism of both phonation onset and eigenmode synchronization, except for one exceptional case.

For the results presented thus far, phonation onset generally occurred due to the synchronization of two closely spaced *in vacuo* eigenvalues. However, in other cases not reported in this study, the eigenvalues which ultimately synchronized were not necessarily closely spaced in the *in vacuo* state. For example, for a slightly different vocal fold geometry, as we varied the flow separation point superiorly along the vocal fold medial surface, the two eigenmodes which synchronized switched from the first and second eigenmodes, to the second and third, and then back to the first and second eigenmodes. Depending on the degree of asymmetry of the flow-induced stiffness matrix, the system sometimes lost stability to a static divergence. Future studies with more comprehensive evaluations over the entire phonatory range are needed to probe the influence of flow parameters (in particular flow separation location) on the movement of the eigenvalues and the factors determining which eigenmodes interact to induce phonation onset.

The linear stability analysis in this study was simplified by a few assumptions which should be relaxed in future studies. First of all, as mentioned in Sec. II, the steady-state problem of the coupled system should be solved for each subglottal pressure. This will give the deformed shape of the vocal folds and the glottal channel based on which the linear stability analysis is performed, and allows the effects of the mean deformation to be investigated. Second, the flow model used in this model was based on one-dimensional potential flow theory and viscous effects were mostly excluded. This means that the conclusions made in this study may not be valid for limiting conditions, such as vanishing structural

damping and an extremely small glottal gap width. The one-dimensional flow description might over simplify the glottal flow dynamics, especially when coupled with a two-dimensional structure model. Other assumptions of the flow such as the fixed flow separation point and zero pressure recovery beyond flow separation should also be relaxed. A more realistic description of the flow, such as the Navier-Stokes equations, should be used instead. The linear stability analysis should then be repeated combining the Navier-Stokes flow with a continuum structure model. On the structural side, a major assumption made in this study was the linear stress-strain relationship of the vocal fold, a condition under which the mean stress and strain in the vocal fold structure can be neglected. When the nonlinearity of the vocal fold material properties is included, the equilibrium stress and strain of the vocal fold structure can no longer be neglected from the linearized equations and therefore have to be solved prior to the linear stability analysis. Of course, a nonlinear stress-strain relationship would also affect the equilibrium geometry of the vocal fold, around which the system equations are linearized.

The analysis of this study can be easily extended to a three-dimensional full vocal folds model so that asymmetric phenomena (asymmetry in the vibration pattern or asymmetry of the glottal flow, e.g., Neubauer *et al.*, 2001, 2007) and three-dimensional effects can be studied. The subglottal and supraglottal acoustics, which are known to be another mechanism of energy transfer, also can be included in the model by using appropriate impedance boundary conditions for the superior and inferior ends of the glottal channel (Sec. II B). These topics will be explored in future studies.

This study shows that the two FSI modes at onset can be expressed as a linear combination of the first four *in vacuo* eigenmodes with reasonable accuracy. The two FSI modes also exhibited a high correlation with the empirical eigenfunctions (EEFs) extracted from medial surface dynamics of the vocal folds (Appendix), indicating that the EEFs could be also expressed as a linear combination of a few *in vacuo* eigenmodes within certain accuracy. These eigenmodes (either the two FSI modes expressed in terms *in vacuo* eigenmodes, or the EEFs) may be used as basis functions for vocal fold modeling. Reduced order models can therefore be obtained for the vocal folds by projecting the governing equations of the vocal folds onto these basis functions. Further, by combining these structural eigenmodes with empirical eigenfunctions of the glottal flow, reduced order models may be developed for a coupled fluid-structure system of phonation (Dowell and Hall, 2001). These reduced order models would allow the system dynamics to be represented by a few dominant eigenmodes, and therefore greatly reduce the computational requirement, which makes it possible for these models to be used in practical applications (e.g., clinic voice diagnosis or prediction of voice surgery outcome).

V. CONCLUSIONS

In this study, we investigated the linear stability of a two-dimensional continuum model of the vocal folds coupled with a potential glottal flow for a fixed flow separa-

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Zhang *et al.*: Physical mechanisms of phonation onset   2293

tion point. We showed that the primary aerodynamic mechanism of phonation onset was the flow-induced stiffness, rather than flow-induced damping. The asymmetric nature of the flow-induced stiffness, caused by nonconservative factors such as flow separation or viscous flow resistance, induced a cross-mode coupling effect. Due to this coupling effect, two *in vacuo* eigenvalues of the vocal fold approximated each other in the complex plane and, at onset, collided with each other to produce an unstable eigenmode. The synchronization of the two eigenmodes provided a net energy transfer from the airflow to the vocal fold structure to overcome the structural dissipation. Only under conditions of negligible structural damping and a restricted set of vocal fold geometries did flow-induced damping become the primary mechanism of phonation onset. However, for moderate to high structural damping, the flow-induced stiffness remained the primary aerodynamic mechanism of phonation onset.

ACKNOWLEDGMENTS

APPENDIX: VOLUME-BASED AND SURFACE-BASED EIGENFUNCTIONS OF VOCAL FOLD VIBRATION

In this study the two spatially and temporally orthogonal FSI modes were extracted based on the vibration dynamics of the entire vocal fold volume. In previous experiments, empirical eigenfunctions (EEFs) were extracted from the observed motion of the vocal fold medial surface. The principal component analysis (PCA) was used in both methods of decomposition, but on data of different domain. The two FSI modes are volume-based eigenfunctions and orthogonal in terms of the whole vocal fold volume, while the EEFs are surface based and orthogonal only along the vocal fold surface. To investigate the possible correlations between these two methods of decomposition, a spatio-temporal dataset of the vocal fold displacement along the vocal fold surface was generated using the critical eigenvector at onset for one case with model parameters specified in Eq. (28) and with a loss factor of 0.1. Empirical eigenfunctions were then extracted from this surface displacement data using the principal component analysis. The first two EEFs were found to be enough to reproduce the whole dynamics, each capturing 87.8% and 12.2% of the total energy, respectively. These compare to the relative energy weights of 76.9% and 23.1% for the FSI-1 and the FSI-2 modes, respectively. Figure 14 compares the first two EEFs (left half of each subfigure) and the two FSI modes (right half in each subfigure). The EEFs were almost the same as the two FSI modes, except for a slight difference between the second EEF and the FSI-2 mode. These differences were further quantified by calculating the surface-based dot product (vector product in Eq. (27) without the M matrix) between the two sets of eigenfunctions. The correlation between the FSI-1 mode and the first EEF was 99.9%, while the correlation between the FSI-2 mode and the second EEF was 98.1%. Although the two sets of eigenfunctions were generated based on different input data, they were

EEF1 and FSI–1   EEF2 and FSI–2
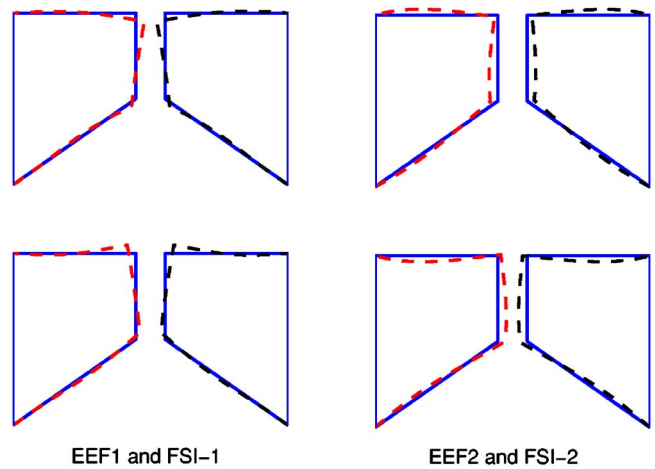
FIG. 14. (Color online) Comparison between the two FSI modes and the first two EEFs of the vibration dynamics along the vocal fold surface. In each subfigure, the left half is the EEF while the right half is the FSI mode. —: equilibrium positions; – –: maximum or minimum displacement of the eigenfunctions superimposed on the equilibrium projections. Model parameters given in Eq. (28), $Q=Q_2\ddot{q}+Q_1\dot{q}+Q_0q$, $\sigma=0.1$.

highly similar in terms of vibration patterns and only slightly different in terms of relative energy weights. This high correlation is probably due to the fact that the vocal fold in this case was nearly incompressible ($\nu=0.47$, Eq. (28)).

Auregan, Y., and Depollier, C. (**1995**). "Snoring: linear stability analysis and *in-vitro* experiments," J. Sound Vib. **188**, 39–54.

Benjamin, T. B. (**1960**). "Effects of a flexible boundary on hydrodynamic stability," J. Fluid Mech. **9**, 513–532.

Benjamin, T. B. (**1963**). "The threefold classification of unstable disturbances in flexible surfaces bounding inviscid flows," J. Fluid Mech. **16**, 436–450.

Berry, D. A., Herzel, H., Titze, I. R., and Krischer, K. (**1994**). "Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions," J. Acoust. Soc. Am. **95**, 3595–3604.

Berry, D. A., Montequin, D. W., and Tayama, N. (**2001**). "High-speed digital imaging of the medial surface of the vocal folds," J. Acoust. Soc. Am. **110**, 2539–2547.

Berry, D. A., and Titze, I. R. (**1996**). "Normal modes in a continuum model of vocal fold tissues," J. Acoust. Soc. Am. **100**, 3345–3354.

Berry, D. A., Zhang, Z., and Neubauer, J. (**2006**). "Mechanisms of irregular vibration in a physical model of the vocal folds," J. Acoust. Soc. Am. **120**(3), EL36–EL42.

Carpenter, P. W., and Garrad, A. D. (**1985**). "The hydrodynamic stability of flow over Kramer-type compliant surfaces. Part 1. Tollmien-Schlichting instabilities," J. Fluid Mech. **155**, 465–510.

Carpenter, P. W., and Garrad, A. D. (**1986**). "The hydrodynamic stability of flow over Kramer-type compliant surfaces. Part 2. Flow-induced surface instabilities," J. Fluid Mech. **170**, 199–232.

Cook, D., and Mongeau, L. (**2007**). "Sensitivity of a continuum vocal fold model to geometric parameters, constraints, and boundary conditions," J. Acoust. Soc. Am. **121**, 2247–2253.

Cullen, J. S., Gilbert, J., and Campbell, D. M. (**2000**). "Brass instruments: Linear stability analysis and experiments with an artificial mouth," Acustica **86**, 704–724.

Deverge, M., Pelorson, X., Vilain, C., Lagree, P. Y., Chentouf, F., Willems, J., and Hirschberg, A. (**2003**). "Influence of collision on the flow through in-vitro rigid models of the vocal folds," J. Acoust. Soc. Am. **114**, 3354–3362.

Dowell, E. H., and Hall, K. C. (**2001**). "Modeling of fluid-structure interaction," Annu. Rev. Fluid Mech. **33**, 445–490.

Duncan, J. H. (**1987**). "A comparison of wave propagation on the surfaces of simple membrane walls and elastic coatings bounded by a fluid flow," J. Sound Vib. **119**, 565–573.

Flanagan, J. L., and Landgraf, L. L. (**1968**). "Self oscillating source for

vocal tract synthesizers," IEEE Trans. Audio Electroacoust. **16**, 57–64.

Grotberg, J. B., and Jensen, O. E. (**2004**). "Biofluid mechanics in flexible tubes," Annu. Rev. Psychol. **36**, 121–147.

Guckenheimer, J., and Holmes, P. (**1983**). *Nonlinear Oscillations, Dynamical Systems, and Bifurcation of Vector Fields* (Springer-Verlag, New York).

Holmes, P. J. (**1977**). "Bifurcations to divergence and flutter in flow-induced oscillations: A finite dimensional analysis," J. Sound Vib. **53**(4), 471–503.

Holmes, P., Lumley, J. L., and Berkooz, G. (**1996**). *Turbulence, Coherent Structures and Symmetry* (Cambridge University Press, Cambridge, UK).

Horacek, J., and Svec, J. G. (**2002**). "Aeroelastic model of vocal-fold-shaped vibrating element for studying the phonation threshold," J. Fluids Struct. **16**(7), 931–955.

Huang, L. (**2001**). "Viscous flutter of a finite elastic membrane in poiseuille flow," J. Fluids Struct. **15**, 1061–1088.

Ishizaka, K. (**1981**). "Equivalent lumped-mass models of vocal fold vibration," in *Vocal Fold Physiology*, edited by K. N. Stevens and M. Hirano (University of Tokyo, Tokyo), pp. 231–244.

Ishizaka, K. (**1988**). "Significance of Kaneko's measurement of natural frequencies of the vocal folds," in *Vocal Physiology: Voice Production, Mechanisms and Functions*, edited by Osamu Fujimara (Raven, New York), pp. 181–190.

Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J. **51**, 1233–1267.

Ishizaka, K., and Matsudaira, M. (**1972**). "Fluid mechanical considerations of vocal cord vibration," *Monogr. 8*, Speech Commun. Res. Lab., Santa Barbara, CA.

Kirillov, O. N. (**2005**). "A theory of the destabilization paradox in non-conservative systems," Acta Mech. **174**, 145–166.

Kounadis, A. N. (**1992**). "On the paradox of the destabilizing effect of damping in non-conservative systems," Int. J. Non-Linear Mech. **27**(4), 597–609.

Landahl, M. T. (**1962**). "On the stability of a laminar incompressible boundary layer over a flexible surface," J. Fluid Mech. **13**, 609–632.

Lighthill, M. J. (**1978**). *Waves in Fluids* (Cambridge University Press, New York).

Lucey, A. D., and Carpenter, P. W. (**1993**). "The hydroelastic stability of three-dimensional disturbances of a finite compliant wall," J. Sound Vib. **165**, 527–552.

Luo, X. Y., and Pedley, T. J. (**1996**). "A numerical simulation of unsteady flow in a two-dimensional collapsible channel," J. Fluid Mech. **314**, 191–225.

Matsuzaki, Y., and Fung, Y. C. (**1977**). "Stability analysis of straight and buckled two-dimensional channels conveying an incompressible flow," J.

Appl. Mech. **44**, 548–552.

Matsuzaki, Y., and Fung, Y. C. (**1979**). "Non-linear stability analysis of a two-dimensional model of an elastic tube conveying a compressible flow," J. Appl. Mech. **46**, 31–36.

Neubauer, J., Mergell, P., Eysholdt, U., and Herzel, H. (**2001**). "Spatio-temporal analysis of irregular vocal fold oscillations: Biphonation due to desynchronization of spatial modes," J. Acoust. Soc. Am. **110**, 3179–3192.

Neubauer, J., Zhang, Z., Miraghaie, R., and Berry, D. A. (**2007**). "Coherent structures of the near field flow in a self-oscillating physical model of the vocal folds," J. Acoust. Soc. Am. **121**, 1102–1118.

O'Reilly, O. M., Malhotra, N. K., and Namachchivaya, N. S. (**1996**). "Some aspects of destabilization in reversible dynamical systems with application to follower forces," Nonlinear Dyn. **10**, 63–87.

Pierucci, M. (**1977**). "Surface waves of an elastic medium in the presence of an inviscid flow field," J. Acoust. Soc. Am. **61**, 965–971.

Steinecke, I., and Herzel, H. (**1995**). "Bifurcations in an asymmetric vocal fold model," J. Acoust. Soc. Am. **97**, 1874–1884.

Story, B. H., and Titze, I. R. (**1995**). "Voice simulation with a body-cover model of the vocal folds," J. Acoust. Soc. Am. **97**, 1249–1260.

Titze, I. R. (**1988**). "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**(4), 1536–1552.

Titze, I. R., Schmidt, S. S., and Titze, M. R. (**1995**). "Phonation threshold pressure in a physical model of the vocal fold mucosa," J. Acoust. Soc. Am. **97**, 3080–3084.

Titze, I. R., and Strong, W. J. (**1975**). "Normal modes in vocal cord tissues," J. Acoust. Soc. Am. **57**, 736–744.

Weaver, D. S. (**1974**). "On the non-conservative nature of "Gyroscopic conservative" systems," J. Sound Vib. **36**, 435–437.

Williamson, C. H. K., and Govardhan, R. (**2004**). "Vortex-induced vibrations," Annu. Rev. Fluid Mech. **36**, 413–455.

Yeo, K. S. (**1988**). "The stability of boundary-layer flow over single- and multi-layer viscoelastic walls," J. Fluid Mech. **196**, 359–408.

Yeo, K. S., (**1992**). "The three-dimensional stability of boundary-layer flow over compliant walls," J. Fluid Mech. **238**, 537–577.

Yeo, K. S., Khoo, B. C., and Chong, W. K. (**1994**). "The linear stability of boundary-layer flow over compliant walls - the effects of the wall mean state, induced by flow loading," J. Fluids Struct. **8**, 529–551.

Zhang, Z., Neubauer, J., and Berry, D. A. (**2006a**). "The influence of subglottal acoustics on laboratory models of phonation," J. Acoust. Soc. Am. **l20**(3), 1558–1569.

Zhang, Z., Neubauer, J., and Berry, D. A. (**2006b**). "Aerodynamically and acoustically driven modes of vibration in a physical model of the vocal folds," J. Acoust. Soc. Am. **120**(5), 2841–2849.

On pressure-frequency relations in the excised larynx

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 Communication Disorders, Bowling Green State University, Bowling Green, Ohio 43403

The purpose of this study was to find relationships between subglottal pressure (P_s) and fundamental frequency (F_0) of phonation in excised larynx models. This included also the relation between F_0 and its rate of change with pressure (dF/dP). Canine larynges were prepared and mounted over a tapered tube that supplied pressurized, heated, and humidified air. Glottal adduction was accomplished either by using two-pronged probes to press the arytenoids together or by passing a suture to simulate lateral cricoarytenoid muscle activation. The pressure-frequency relation was obtained through a series of pressure-flow sweep experiments that were conducted for eight excised canine larynges. It was found that, at set adduction and elongation levels, the pressure-frequency relation is nonlinear, and is highly influenced by the adduction and elongation. The results indicated that for the lower phonation mode, the average rate of change of frequency with pressure was 2.9 ± 0.7 Hz/cm H_2O, and for the higher mode was 5.3 ± 0.5 Hz/cm H_2O for adduction changes and 8.2 ± 4.4 Hz/cm H_2O for elongation changes. The results suggest that during speech and singing, the dF/dP relationships are taken into account. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2772230]

I. INTRODUCTION

Subglottal pressure (P_s) and fundamental frequency (F_0) of the voice are two primary control variables of phonation, the former highly related to the loudness of the sound produced, the latter to the perceived pitch of the voice. These two variables are not independent. Subglottal pressure has been shown to have an influence on the fundamental frequency (Lieberman *et al.*, 1969; Baer, 1979; Rothenberg and Mahshie, 1986, Titze, 1989) and intensity of phonation (e.g., Koyama *et al.*, 1971). A derived parameter that is obtained from the pressure-frequency relation is the slope of that relation, or rate of change in F_0 with pressure, dF/dP. This value would be constant at any frequency if F_0 versus P_s were linear. Titze's detailed empirical-theoretical approach using excised canine larynges and prior empirical relationships among phonatory variables (Titze, 1989) suggests a reduction in dF/dP as frequency increases in the modal register. However, the parameter dF/dP is reported to have almost a "V" shape behavior as a function of frequency in human phonation (Tanaka *et al.*, 1997; Kataoka and Kitajima, 2001; Kataoka *et al.*, 2001). That is, it has a minimum at some frequency and increases as frequency increases or decreases from that value. The left-hand side of the "V-shaped" structure of the dF/dP versus F_0 function is consistent with Titze's empirical-theoretical study. The right-hand side of the V-shaped may be explained by a reduction in mass or reduction in vibratory depth at higher frequencies (Kataoka *et al.*, 2001; Tanaka *et al.*, 1997).

The decrease of dF/dP with F_0 is also seen with live dog experimentation where cricothyroid (CT) muscle activation is used to elongate the vocal folds, as shown in the work of Hsiao *et al.* (2001). They examined three dogs at three CT activation levels. For each condition, F_0 increased with P_s increase, and linear fitting was used. The slopes of those linear fits decreased as F_0 increased (that is, with greater CT stimulation).

Baer (1979) showed that typical values of dF/dP were acquired when using the technique of producing sudden increases in subglottal pressure (pushing on the chest or abdomen) during the approximately 30 ms prior to the laryngeal response, which might change vocal fold length and adduction.

Atkinson (1978) studied a single subject with EMG and recorded P_s during the production of English statements and questions. In general, the CT muscle was the primary factor that related to F_0 production. However, his correlation analyses suggested that P_s was a dominant control factor in statements but not in questions, where F_0 followed P_s in the former, but in the latter, there was a strong necessity to create fast F_0 rise by using the CT muscles (as well as the lateral cricoarytenoid [LCA] and vocalis muscles). Furthermore, at the lowest speaking pitch (80–100 Hz for the single male subject), CT was not a factor and P_s controlled F_0 (with minor help from the LCA). At the highest speaking pitches (140–160 Hz for the subject), CT appears to have been the dominant controlling factor, and P_s appears not to have contributed (indeed, P_s decreased with F_0, suggesting a synergistic mechanism for more than F_0 control). This study by Atkinson suggests, then, that P_s appears to participate in the

control of F_0 at lower pitches and in linguistic statements. It is noted that the study examined statements and questions, not sustained vowel phonations.

In another EMG study of CT muscle with F_0 and P_s recordings by Collier (1975), Dutch sentences were read with various stress. The CT activity corresponded well to the direction, magnitude, and rate of change of F_0, confirming that CT appears to be the dominant F_0 controlling factor. However, P_s appeared to play a significant role in controlling F_0 declination (by P_s reduction). It is noted that the P_s traces in the figures of Collier's study showed the same number of local maxima as the CT and F_0 did, but the timing of the P_s peaks were not consistent, suggesting that the role of P_s is more complicated than simply indicating that it helps to control F_0 declination.

The values for dF/dP cover a wide range across different studies. For example, Ladefoged and McKinney (1963) recorded /a/ vowel prolongations using a variety of pitches and loudness levels while recording subglottal pressure via the esophageal balloon technique and volume flow with a pneumatic mask system. Their Fig. 4 suggests an increase of approximately 6 Hz per cm H_2O. Lieberman *et al.* (1969) studied the change of F_0 with respect to the change of transglottal pressure via sinusoidal alterations of the air pressure within the oral cavity (and therefore the supraglottal airway) during prolonged vowels. They measured a range of approximately 3–18 Hz/cm H_2O, with higher values for higher F_0 (within the modal register for one subject). Ohala (1978) demonstrated that F_0 decreases when the sternohyoid muscle contraction increases. When the sternohyoid muscle contracts, the larynx lowers, thus providing a condition in which the vocal fold tissues may be more pliable. In contrast, if the larynx rises, the vocal folds rise, and the tracheal tissue would create a downward pull that stretches the vocal fold tissue in the vertical direction. This is supported by Ohala's observations that high vowels and high pitches show a larger laryngeal ventricle, suggesting that there is a vertical pull on the false vocal folds but a downward pull on the true vocal folds. Kori *et al.* (1990) also show that the sternohyoid muscle is an active control of pitch-lowering in Japanese. Strik and Boves (1995) are convincing in their assessment that the general declination of F_0 for statements (and increase for questions) is highly dependent upon the change of P_s over the utterance, once the influence of the laryngeal muscles is removed. They applied a statistical regression to their data relating F_0 to P_s, vocalis muscle activity, and sternohyoid muscle activity. They are quick to indicate that the decision of relevance of variables to the control of F_0 depends on how F_0 declination is defined.

In these prior studies, mean pressure and flow were typically varied in a stepwise fashion with few pressure-frequency data points, and the derivative of the pressure-frequency relation was calculated from those stepwise differences. The purpose of the study reported here was to provide more continuous pressure-frequency data, investigate the range and structure of the parameter dF/dP during continuous phonation and for transitions in phonation mode, and enhance the theoretical approach to this phenomenon.

TABLE I. Canine larynges used in this study and information on their size. Vocal fold length was measure at rest position from the anterior commissure to the tip of vocal processes.

Larynx	Weight (kg)	Gender	VFL (mm)	Major controls
CL24	25	M	13	Flow rate, adduction
CL27	20	M	13	Flow rate, adduction
CL28	25	N/A	11	Flow rate, adduction
CL31	21	M	15	Flow rate, tension
CL33	22	M	16	Flow rate, tension
CL42	21	F	13	Flow rate, adduction, tension
CL52	17	M	12	Flow rate, adduction, tension
CL64	18	M	13	Flow rate, adduction, tension

II. METHODS

In this study, eight excised canine larynges (see Table I) were used. It is the case with such studies that the tension of the vocal fold tissues is passive because there is no contraction of the thyroarytenoid muscles. The results can be thought, therefore, to be limited because of this lack of complete physiology; however, a study of the intricate mechanisms of phonation should begin with a passive system as a basis for consideration of more complicated studies. Moreover, it will be shown that the results are similar to those published using human subjects, suggesting that much of the control of frequency via subglottal pressure may indeed have a significant passive aspect. Furthermore, a study of the effects of adduction per se can be performed with a passive system for, again, baseline effects. In this regard it is pointed out that the study deals with the "passive" vocal fold approximation through arytenoid cartilage adjustment with mimicked interarytenoid, lateral cricoarytenoid, and lateral thyroarytenoid muscle contraction, the effect of which is to bring the vocal processes closer together. The other way to bring the membranous vocal folds toward each other is through bulging of the medial vocal fold surface through vocalis contraction; this method is not explored here.

The excised canine larynx has been used as a model for human phonation for many years. Although the vocal fold differs from the human in that there are only two layers of the lamina propria instead of three (the canine has no vocal ligament), and the lamina propria is deeper than for the human (about 3 mm for the canine compared to about 1.1 mm for the human), the similarities in size, vocal fold length, vibratory motion, frequency and subglottal pressure ranges, flow rate, ease of obtaining phonation during empirical bench work, and controllability and durability for such work make the excised canine larynx highly appropriate for studying passive aspects of phonation (including basic kinematic, aerodynamic, and acoustic characteristics of phonation). The excised larynx model usually operates at higher flow rates than for reported line for human phonation (Alipour *et al.*, 2001).

Excised canine larynges were obtained following cardiovascular research experiments at the University of Iowa Hospitals and Clinics. They were quick-frozen using liquid nitrogen for storage and slowly thawed in a refrigerator prior to use. Each larynx was mounted on $\frac{3}{4}$ in. PVC tubing mimick-
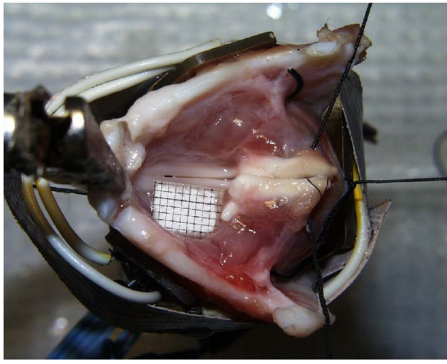
FIG. 1. (Color online) Mounted excised canine larynx with adduction and elongation sutures and EGG electrodes. The EGG electrodes (at the top and bottom) are attached with pins to the thyroid cartilage and secured with duct tape. The shown millimeter grid was used for length calibration and was removed prior to the runs.

ing the trachea so that the glottis was easily viewed by a camera and accessible by the equipment. Adduction was created either by the pressing of arytenoid cartilages together with multipinned prongs (simulating IA muscle contraction) while metal shims of various thickness (0.1–1.0 mm) were placed between them, or by a pair of sutures pulling on the muscular process of each arytenoid cartilage to simulate lateral cricoarytenoid and (lateral) thyroarytenoid muscle action as in arytenoid adduction. Different adduction methods were performed in separate larynges.

To achieve the arytenoid adduction using sutures, the muscular processes of the arytenoids were identified and a threaded curved needle was passed through the muscular process, a knot was made, and the needle removed. Then the free end of the suture line was threaded through a straight needle and directed anterio-laterally and pulled out from the anterior side, through the thyroid lamina. During the experiment when this suture was pulled by the addition of graded weights, the muscular process was rotated anterior-medially, bringing the vocal processes toward each other. These weights ranged from approximately 100 to 200 g for low to high adductory condition in the canine larynx. It is noted that the actual value of the weights was of little concern, but the effect of the weights in grading the level of adduction was. That is, different specimens required different amounts of weight to move the cartilages relative to each other to change the adduction (or length), and these varied according to individual differences across the larynges.

Elongation and corresponding tension of the vocal folds were simulated either by pulling the anterior side of the thyroid cartilage with a micrometer-controlled alligator clip attached to the middle of the thyroid cartilage, or pulling the arytenoids posteriorly by two bilateral sutures (attached at medial locations on both arytenoids) via graded weights. Both methods provided elongation and tension control of the vocal folds.

Figure 1 shows a mounted excised larynx with adduction and elongation sutures. Air from a pseudolung passed through a heater and humidifier (ConchaTherm® unit, RCI Laboratories) to achieve 37 °C and 100% humidification and entered the larynx via tapered $\frac{3}{4}$ in. tubing. Electrode plates

from a Synchrovoice electroglottograph (EGG) were placed on the thyroid laminae to obtain the EGG signal during phonation. The EGG signal was used to extract fundamental frequency. The time-varying subglottal pressure was recorded using a pressure transducer (Microswitch 136PC01G1, with an approximate bandwidth of 0–1 kHz) 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) and measured with a flow pneumotach (Rudolph 4700 with Validyne DP103 pressure transducer), both devices located upstream of the humidifier. During the experiments, the top view of the larynx was recorded onto videotape for later observation.

The experiments started with at least two pressure-flow sweeps for each condition. These sweeps consisted of a gradual increase in pressure and flow (upward sweep) from the lowest level needed to initiate phonation to the highest level at which phonation could be maintained. This was performed manually by increasing the flow rate using a rotary valve such that the flow increase was gradual and smooth. Each sweep lasted about 20 s and was used to identify the ranges of sustained oscillations in the pressure-flow-frequency domain, and to reveal any region with changes in the mode of phonation, with adduction and elongation as the major prephonatory control parameters. The second sweep was started from the highest flow rate and decreased (downward sweep) until oscillation stopped, essentially the reverse of the first sweep.

The EGG, subglottal pressure, and mean flow rate signals were recorded into a computer using an A/D (14 bit with dc coupling) board and software (DATAQ Instruments). These signals were converted to physical quantities in MAT-LAB and used for the aerodynamic and acoustic analyses. The signals were plotted first and the phonatory section was selected for the analysis. The selected sweep section was divided into 50–100 segments such that each segment included typically 10–20 phonatory cycles. After identifying the fundamental frequency range from the spectrogram, the EGG signal was digitally low-pass filtered just above the maximum F_0 (cutoff $\sim1.5*F_0$, that is, 300–500 Hz) and used to calculate the average period of each of the segments (using the zero crossing technique). The mean subglottal pressure and mean flow rate for each segment were also obtained. Glottal flow resistance was calculated from the slope of the mean pressure-flow data (flow resistance has been found to be highly linear for excised canine larynges; see Alipour *et al.*, 1997).

The frequency-pressure derivative dF/dP was calculated from the pressure-frequency data in the following manner. For most phonatory conditions without a mode change, the pressure-frequency data were monotonic and a third-order polynomial fit was a convenient choice that provided a continuous function. However, a third-order polynomial may miscalculate (overestimate) the slopes in certain regions of the original data such as end points. Thus, the range for the frequency values was divided into eight to ten linear seg-

TABLE II. Effects of adduction on the pressure-frequency relation in canine excised larynges (at rest length, no tension). The first column identifies the larynges and the second column indicates the case of the larynx followed by a dash and then either U or D for upward sweep and downward sweep, respectfully. Columns 7-9 are for the second oscillation mode (if it existed). The dF/dP is in Hz/cm-H$_2$O. The flow resistance is in cm-H$_2$O/(l/s). GFR stands for glottal flow resistance.

Larynx	Case	ADD	F_0 range	dF/dP Lo, Hi, Ave	GFR	F_0 range	dF/dP Lo, Hi, Ave	GFR
CL24	**A-U**	Low	122–142	1.0, 2.9, 2.3	21.5			
CL24	**A-D**	Low	119–140	1.5, 3.6, 2.4	21.8			
CL24	**C-U**	Med	128–156	1.4, 2.2, 1.9	26.0			
CL24	**C-D**	Med	128–156	1.4, 2.6, 2.0	27.7			
CL24	**D-U**	High	139–161	1.3, 2.9, 1.9	30.9			
CL24	**D-D**	High	131–159	1.3, 2.6, 2.1	32.2			
CL27	**K-U**	Low	104–116	1.1, 6.9, 3.0	29.8	169–193	3.4, 11.8, 5.6	18.5
CL27	**K-D**	Low	83–97	1.1, 6.4, 2.7	25.0	145–197	4.1, 6.9, 4.6	16.2
CL27	**L-U**	Med	108–119	1.0, 9.0, 2.4	21.5	172–214	3.3, 12.0, 9.0	19.2
CL27	**L-D**	Med	85–95	1.3, 5.7, 2.4	24.0	143–204	4.2, 6.8, 5.1	18.0
CL27	**M-U**	High	105–125	2.5, 11.2, 3.5	22.7	180–219	5.3, 9.1, 6.5	17.9
CL27	**M-D**	High	91–97	4.5, 5.5, 5.1	29.0	149–217	4.0, 5.3, 4.9	18.2
CL28	**A-U**	Low	232–250	1.1, 2.4, 2.0	20.3			
CL28	**A-D**	Low	209–243	1.6, 4.5, 3.4	29.8			
CL28	**B-U**	Med	246–284	2.5, 5.2, 4.2	23.1			
CL28	**B-D**	Med	217–286	2.6, 5.2, 3.5	31.0			
CL28	**C-U**	High	241–280	2.4, 7.4, 3.8	31.8			
CL28	**C-D**	High	239–294	3.0, 6.9, 5.1	33.5			
CL42	**D-U**	Low	75–130	2.1, 8.6, 4.0	20.2			
CL42	**D-D**	Low	80–134	2.2, 6.8, 3.8	21.8			
CL42	**E-U**	Med	76–134	2.3, 8.4, 4.0	25.4			
CL42	**E-D**	Med	80–135	2.1, 7.0, 4.0	25.2			
CL42	**F-U**	High	97–120	2.3, 12.2, 6.9	25.0	262–276	5.9, 9.7, 6.7	25.3
CL42	**F-D**	High	82–92	2.3, 3.5, 2.5	26.2	236–273	2.7, 10.0, 5.5	25.8
CL64	**B-U**	Low	170–260	0.3, 9.3, 6.3	34.4			
CL64	**B-D**	Low	150–250	2.5, 9.6, 5.1	30.6			
CL64	**C-U**	Med	150–250	2.4, 8.5, 6.7	34.7			
CL64	**C-D**	Med	150–255	2.7, 10.4, 5.7	34.5			
CL64	**E-U**	High	161–196	1.3, 9.0, 3.4	38.4			
CL64	**E-D**	High	213–261	3.4, 8.9, 6.5	42.2			

ments and their linear slopes were used in the dF/dP function estimation. However, the derivative dF/dP obtained in this manner was not a continuous function. For the phonatory conditions with a mode change, modeling was with linear regressions, but only the sections before and after the mode changes were so modeled.

The pressure-frequency relations were typically nonlinear and the numerical derivative of the pressure-frequency data usually included numerical noise or amplified noise due to the jitter in the vibration of the vocal folds. A numerical smoothing or digital filtering would have reduced this noise if the number of data points was large enough and the points were equally spaced. Also, for all eight larynges, the glottal flow resistance and dF/dP were calculated both in upward sweeps and downward sweeps with a linear regression model and reported in Tables II and III.

III. RESULTS

Figure 2 shows a portion of the pressure and flow signals at the start of an oscillation. For this case, the pressure increased and the mean flow rate decreased initially (also evident from their low-pass filtered signals, thick lines). As the oscillation built up, the peak-to-peak values of the pressure and flow signals increased. The EGG signal (first trace) usually grew more slowly than the other two (although this observation is not seen in Fig. 2, all three signals were "steady" by about 60 ms) and was sensitive to the oscillation condition and the mounting of the electrodes. These traces illustrate the broadband frequency of the recorded signals.

Figure 3 includes four curves that are the mean values of consecutive segments of the signals (see above) for excised larynx CL33 in an upward sweep as a function of time. The curves are, from top to bottom, subglottal pressure, flow rate, fundamental frequency, and subglottal ac pressure amplitude (one half of the peak-to-peak pressure), respectively. For the first 9.4 s, the increase in flow was accompanied by a nearly linear increase in pressure and increase in frequency. During the pressure-flow sweep it is possible for the larynx to shift from one mode of phonation to another, as shown in this example in Fig. 3 at around 9.5 s, where a sudden drop in the fundamental frequency and flow rate was accompanied by a sudden jump in subglottal pressure and pressure amplitude. The corresponding changes in the pressure and flow signals indicate a change in the glottal flow resistance after the mode

TABLE III. Effects of elongation or tension on the pressure-frequency relation in canine excised larynges (at low adduction levels). The description of the columns is similar to Table I. The range of elongation (for larynxes CL31, CL33, and CL42) from low to high was 2 mm. The range of tension applied to the thyroid cartilage to elongate the folds for larynx CL52 ranged from 132 g (low) to 300 g (high). The missing data for larynx CL42 in columns 4–6 was due to the absence of oscillation for the lower mode for medium and higher elongations.

Larynx	Case	TE	F_0 range	dF/dP Lo, Hi, Ave	GFR	F_0 range	dF/dP Lo, Hi, Ave	GFR
CL31	**I-U**	Low	254–297	0, 5.5, 3.9	26.1			
CL31	**I-D**	Low	242–314	2.9, 12.1, 5.2	24.3			
CL31	**J-U**	Med	252–302	2.0, 4.5, 3.9	21.8			
CL31	**J-D**	Med	236–307	2.7, 5.4, 3.2	22.7			
CL31	**K-U**	High	320–350	0, 5.5, 1.6	24.5			
CL31	**K-D**	High	254–341	1.6, 6.2, 5.5	25.7			
CL33	**G-U**	Low	130–140	0.0, 2.0, 1.2	25.3	270–295	2.5, 4.8, 3.7	16.6
CL33	**G-D**	Low	135–142	0.8, 1.2, 1.0	16.7	284–300	1.3, 4.8, 4.0	21.9
CL33	**H-U**	Med	170–180	0.6, 0.8, 0.6	26.1	343–480	2.7, 11.9, 7.8	27.4
CL33	**H-D**	Med				396–465	2.5, 18.6, 5.4	25.4
CL33	**I-U**	High				459–469	3.9, 6.4, 5.1	26.5
CL33	**I-D**	High				456–476	3.7, 6.4, 4.1	26.8
CL42	**O-U**	Low	109–128	2.1, 9.8, 3.6	21.2	268–280	1.3, 6.9, 4.1	20.2
CL42	**O-D**	Low	176–186	0.5, 3.8, 2.2	29.1	232–242	12.4, 12.6, 12.5	19.0
CL42	**T-U**	Med				243–275	3.1, 13.6, 5.1	15.6
CL42	**T-D**	Med				244–294	1.2, 14.5, 7.7	14.6
CL42	**U-U**	High				260–304	7.5, 14.9, 13.6	16.7
CL42	**U-D**	High				239–325	1.6, 13.0, 10.8	17.5
CL52	**N-U**	Low	282–303	0.2, 3.4, 2.9	59.4			
CL52	**N-D**	Low	292–307	1.0, 2.1, 1.2	35.2			
CL52	**O-U**	Med	299–307	1.2, 2.0, 1.4	56.8	314–340	9.9, 13.8, 11.8	37.0
CL52	**O-D**	Med	304–308	0.0, 1.4, 0.6	41.2			
CL52	**P-U**	High	351–364	0.0, 7.7, 6.3	49.3			
CL52	**P-D**	High	357–364	1.2, 9.5, 2.7	59.4			
CL64	**G-U**	Low	165–201	4.2, 4.8, 4.5	33.1			
CL64	**G-D**	Low	247–253	2.2, 4.0, 3.1	26.5	329–346	1.3, 11.1, 6.3	62.2
CL64	**I-U**	Med	184–238	4.6, 8.9, 5.5	32.5			
CL64	**I-D**	Med	198–239	0.8, 4.9, 4.1	32.4			
CL64	**J-U**	High				348–374	5.2, 10.3, 7.7	32.7
CL64	**J-D**	High				348–380	1.3, 5.8, 4.9	33.5

change. For this case, with the decrease in frequency, both of the pressures and the glottal flow resistance were increased. Upon visually inspecting the case shown in Fig. 3, the shift in phonation appeared to be due to entrainment of the lower glottis when the amplitude of motion was sufficiently great (note that vocal fold length was not changed throughout the sweep, just the flow and pressure). Here the two oscillating regions cannot be modeled together due to the pressure and frequency jumps and need to be treated separately.

A. Effects of adduction on pressure-frequency relations

The pressure-frequency relation is nonlinear, as shown in Fig. 4 for the excised larynx CL28. Figure 4 suggests a monotonic relation between F_0 and subglottal pressure, with (in this case) a shift of the curves to the right with greater adduction (suggesting slightly lower F_0 with greater adduction at the same transglottal pressure). The pressure-frequency data points (closed symbols) were modeled with a multisegment linear model to calculate their slope functions. Since the segments are selected by the number of data points, they are not equally spaced in pressure as seen on the graph.

The dF/dP (open symbols) with values on the second Y axis takes on a shape with maximum and minimum locations. The low adduction case (open triangles) shows a minimum around 28–30 cm H_2O pressure or about 250 Hz. The slopes start at around 6.8 Hz/cm H_2O, and drop to a minimum of about 2.5 Hz/cm H_2O at a pressure of about 32 cm H_2O and an F_0 value of approximately 250 Hz. This is similar to the expected (literature-supported) concave "V" shape for the low level of glottal adduction. The higher adduction cases have minimum values in a similar range of subglottal pressure, that is, within approximately 28–32 cm H_2O. There is also a maximum slope for all three curves around 35–38 cm H_2O. Thus, for this case, adduction shifted the frequency relative to subglottal pressure, but the general trend of the frequency-pressure curves was similar relative to gross locations of minimal and maximal dF/dP values, namely, in the middle and near the end of the upward sweeps, respectively.

The curvature pattern and degree of nonlinearity can change from larynx to larynx, and be different within a larynx for different adductions or elongations. For example, the pressure-frequency curves for another excised larynx (CL64) are shown in Fig. 5. Here the degree of nonlinearity and the
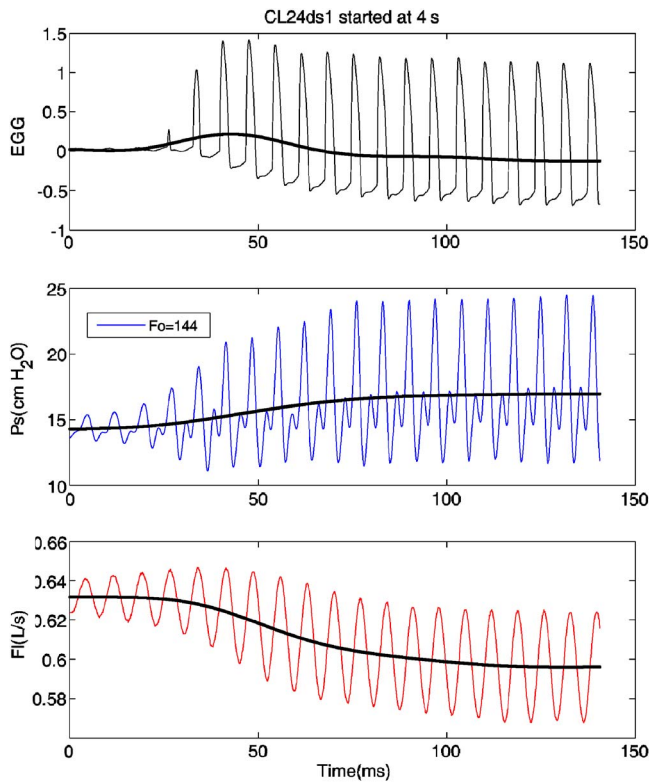
FIG. 2. (Color online) Onset cycles of the EGG, pressure, and flow signals in a pressure-flow sweep for canine larynx CL24. EGG cycles appear with a short delay. The smooth lines indicate average values (low-pass filtered signals).

FIG. 3. (Color online) Time variations of pressure, flow rate, fundamental frequency, and subglottal pressure amplitude for excised larynx CL33. Each point is an average over approximately 10 phonatory cycles. There is a frequency and flow drop between 9 and 10 s and a jump in the pressure signals corresponding to a mode change.

slope of the curves are seen to vary as adduction level increased. In this case, pressure-frequency curves relative to adduction approach each other at pressures around 21–23 cm H_2O. Below these pressures, higher adduction corresponded to lower frequency for the same pressure values. For low adduction, the pressure-frequency curve was more linear at low pressures. The shape of the low adduction dF/dP curves was similar to Fig. 4, but at high adduction, due to the complex shape of the pressure-frequency curve, there was no evident minimum, but all three showed maximum dF/dP values around 20–25 cm H_2O. In general the low-pressure and high-pressure portions of these curves have different behaviors, but because they are monotonic, a linear regression fit can provide an average estimate of their derivatives.

Figure 6 illustrates a different pattern of F_0 change with pressure increase. Here, in the excised larynx CL27 with various adduction levels, as the pressure was increased during a sweep, the frequency had a short rising segment and then flipped to a lower mode. The frequency continued to rise for about 3–5 s, and then jumped to a higher mode. The individual segments of oscillations excluding the transients can be modeled separately. For example, the average of the three slopes of the lower modes (from linear approximation) is about 2.5 Hz/cm H_2O and for the higher mode is about 5.5 Hz/cm H_2O. The mechanism of the mode changes is beyond the scope of this work and will be discussed in future work.

Table II presents the frequency and dF/dP ranges for five excised larynges for various adduction levels, both in upward and downward sweeps. In the first and second columns the larynx ID, case, and direction of the sweeps, upward (U) and downward (D), are presented. The third column shows adduction level and the fourth column shows the
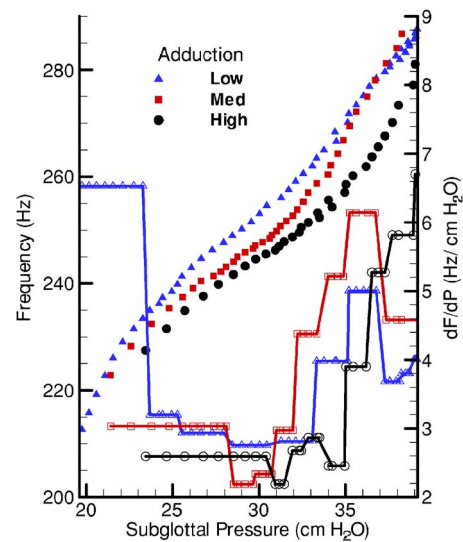
FIG. 4. (Color online) Pressure-frequency relations in canine larynx CL28 during an ascending sweep with three levels of adduction via suturing at the vocal processes. Closed symbols represent frequency data, and open symbols represent dF/dP calculated from the multisegment linear model with values on the right.
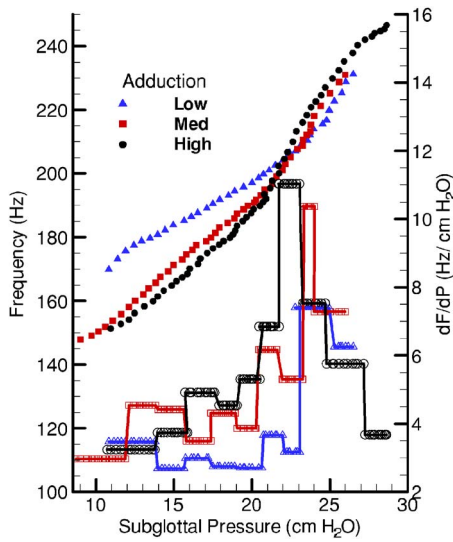
FIG. 5. (Color online) Pressure-frequency relations in canine larynx CL64 during an ascending sweep with three adduction levels. Symbols have the same definition as in Fig. 4.

range of frequency during stable oscillations. In column five, the lowest (Lo) and the highest (Hi) values of dF/dP for that frequency range, along with the average slope, were calculated from a linear model for that range. The next column is the glottal flow resistance (GFR). Similar information for the cases with mode changes is given in the last three columns. The downward sweeps show only minor differences in frequency ranges, dF/dP slopes, and flow resistance compared to the upward sweep data. The average dF/dP for the first phonation mode was 2.9 ± 0.7 Hz/cm H_2O and for the second (higher frequency) mode was 5.3 ± 0.5 Hz/cm H_2O. The average glottal flow resistance for the first mode was 26.1 ± 4.1 cm $H_2O/(l/s)$ and for the second mode 19.9 ± 3.6 cm $H_2O/(l/s)$ (statistically significant difference,
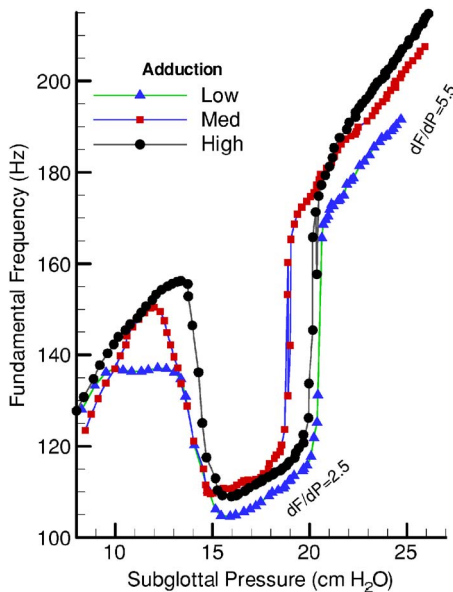
FIG. 6. (Color online) Pressure-frequency relations in canine larynx CL27 during an ascending sweep with mode changes at three adduction levels (low, medium, and high). The indicated slopes are average from linear model for that segment.
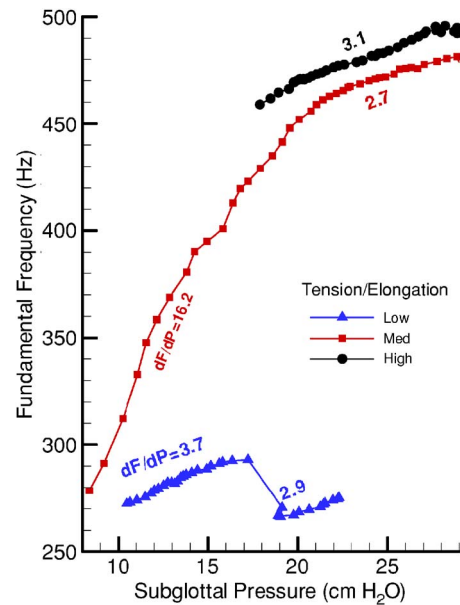
FIG. 7. (Color online) Pressure-frequency relations in canine larynx CL33 during an ascending sweep with three levels of elongation and averaged segment slopes.

$p=0.0007$, using Student t-test). Thus the second mode had higher dF/dP and lower glottal flow resistance values.

B. Effects of tension on pressure-frequency relations

The effects of elongation and tension on pressure-frequency relations are presented in Figs. 7–9. These figures indicate the nonlinearity that may be enhanced due to the tension and elongation. The average slopes from linear models are also indicated on the segments. Figure 7 shows the results of elongation for the excised larynx CL33 at three levels of elongation: low (0–1 mm), medium (1–2 mm), and high (2–4 mm). The relative arrangement of the three curves is reminiscent of larynx CL28 shown in Fig. 4. The
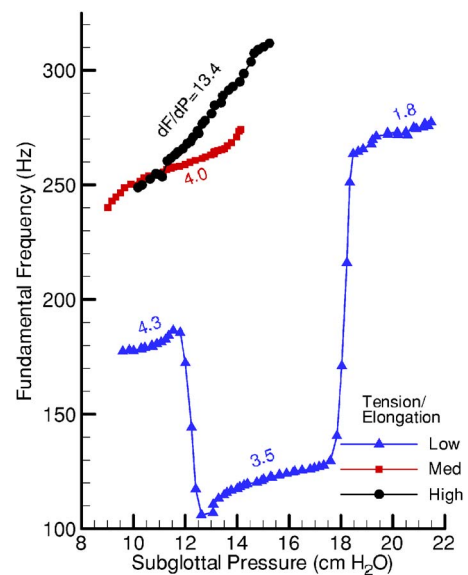
FIG. 8. (Color online) Pressure-frequency relations in canine larynx CL42 during an ascending sweep with three levels of elongation and averaged segment slopes.

2302   J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

F. Alipour and R. C. Scherer: Pressure-frequency relations
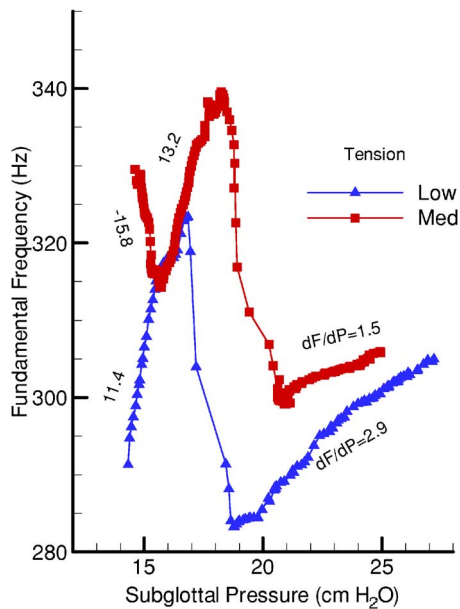
FIG. 9. (Color online) Pressure-frequency relations in canine larynx CL52 during an ascending sweep with mode changes at two tension levels (low and medium).

tension experiments on this larynx not only resulted in steep curves with slopes as high as 16.2 Hz/cm H_2O for the medium elongation, but also they induced a mode change that created a discontinuity in pressure, flow rate, and other glottal parameters for the low elongation. The low-elongation curve is relatively linear with a pitch drop at around 18 cm H_2O and an average slope of 3.7 Hz/cm H_2O for the first segment and 2.9 Hz/cm H_2O for the second segment. Similarly, the results of Fig. 8 for the excised larynx CL42 indicate high nonlinearity and pitch jumps with the corresponding local average slopes indicated on the various segments.

In another larynx (CL52), the vocal folds were subjected to different degrees of longitudinal tension with arytenoid pulls and then the pressure-flow sweep method was performed, as shown in Fig. 9. For a low-tension value (85 g), the frequency-pressure curve had two increasing branches with a discontinuity ending at approximately 18 cm H_2O. For a higher tension value (130 g), more complicated F_0 patterns occurred as P_s was increased. Observations of the video recordings of the larynx suggest that the F_0 *decreases* in these patterns were due to increasing the amount of tissue mass in vibration as the pressure increased, whereas the *increases in F_0* were due to the expected cause of greater tension with greater excursion due to pressure increase. The corresponding local average slopes are shown in Fig. 9.

Table III shows the frequency ranges and the average values of dF/dP and glottal flow resistance for five excised larynges at various elongation and tension levels, both in upward and downward sweeps. Two of the larynges (CL42 and CL64) were used for both adduction (Table II) and elongation studies. The effects of tension are primarily observed in the increase of frequencies within the first and second modes. The average dF/dP for the first phonation mode was 2.7 ± 1.2 Hz/cm H_2O and for the second (higher frequency) mode 8.2 ± 4.4 Hz/cm H_2O. The average glottal flow resis-

tance for the first mode was 34.2 ± 13.1 cm $H_2O/(l/s)$ and for the second mode 22.9 ± 7.9 cm $H_2O/(l/s)$ (statistically significant difference: the test variable is glottal resistance, with 1 degree of freedom; $p=0.012$, using Student t-test). Similar to the adduction effects, the second mode of tension cases had higher dF/dP and lower glottal flow resistance. The overall glottal flow resistance and dF/dP are higher than for the adduction conditions due to the added longitudinal tensions. The rate of frequency increase relative to subglottal pressure from linear modeling calculationsreached as high as 16.1 for the second mode. However, the first mode of phonation had a similar dF/dP to the adduction cases.

IV. Discussion

The results of Strik and Boves (1995) indicated that the regression analysis allowed dF/dP values to fall typically into the $2-7$ Hz/cm H_2O range, with the general trend for F_0 change to follow P_s closely. This was a strong statement of the linguistic importance of P_s to F_0 change in speech. One conclusion that can be drawn from the work by Strik and Boves is that, depending on how F_0 declination is defined, P_s accounts for all of the F_0 general declination over the length of an utterance.

The excised canine data of the study here suggest even a greater complexity of active control of F_0 through P_s change. The nonlinearity of the pressure-frequency relations and dF/dP vs F_0 curves indicate the potential for a relatively large range of dF/dP values, $1-20$ Hz/cm H_2O, greater than the general finding of $2-7$ Hz/cm H_2O. The larger dF/dP values mean a greater F_0 change for a specific change in P_s. If the findings of this study are similar in humans, that dF/dP is highly dependent upon F_0, speakers continually would either be *correcting for or taking advantage of* this change with F_0 in order to express F_0 during speech and singing. A similar situation perhaps is the continual counter control of respiratory muscle action over the length of an utterance due reduction of respiratory recoil. Here, we must take into consideration the prevailing dF/dP at the desired F_0 at which we wish to phonate at the moment. Thus, if an utterance has a linguistically based F_0 contour, that contour is controlled by an F_0 dependent dF/dP factor as well as by control of muscles such as the cricothyroid (CT), thyroarytenoid (TA), and sternohyoideus (SH) muscles.

The adduction and elongation for different larynges showed different dF/dP behavior, mostly nonlinear. The results given in Tables II and III showed that for the first mode of oscillation, the rate of change of frequency with pressure reflected dF/dP averages of 2.9 ± 0.7 Hz/cm H_2O, but for the second mode this rate increased to 5.3 ± 0.5 Hz/cm H_2O for adduction changes and 8.2 ± 4.4 Hz/cm H_2O for elongation changes. Thus one possibility for greater control over pitch might be a mechanism that triggers oscillation to the second mode, either by higher adduction or elongation, or both. When these two effects are combined, one might expect a higher dynamic range for the pitch relative to pressure. The significance of this possibility will be studied in future work. It is noted that the effect of higher dF/dP values for
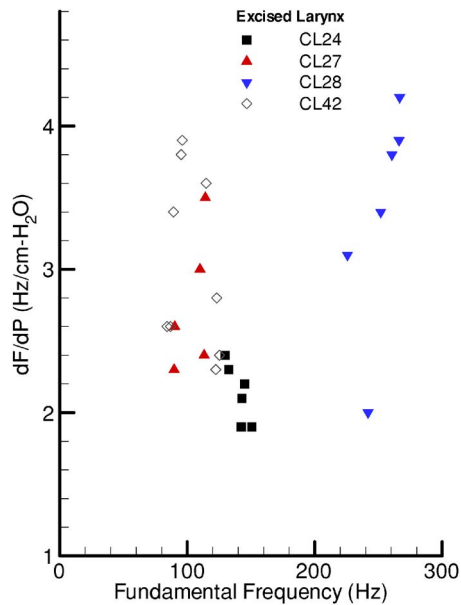
FIG. 10. (Color online) Generalized relationship between dF/dP and F_0 for four larynges of Table II at low, medium, and high adduction levels.

FIG. 11. (Color online) Generalized relationship between dF/dP and F_0 for five larynges of Table III at low, medium, and high tension levels.

higher frequencies is perceptually countered somewhat if one takes the logarithmic nature of pitch into consideration, as in the use of semitones.

When the pressure-frequency relation was a monotonic function, the shape of dF/dP versus pressure had possible maxima and minima. The nonlinearity of the pressure-frequency data suggests that the pressure control of fundamental frequency varies in its sensitivity to F_0 change, depending upon the level of pressure, and less so on the level of adduction. Figure 10 shows the relation between average dF/dP and average F_0 of the oscillation range for four larynges of Table II. The generalized relationship between dF/dP and frequency is similar to Fig. 4 for larynx CL28, and independent of the adduction level, but dF/dP is shown to be dependent on the prevailing F_0 value. The finding that dF/dP is higher at high frequencies can be explained by the effects of pressure on the tension of the vocal folds. As pressure pushes the vocal fold apart, the amount of tensile stress that is experienced by the vocal folds depends on the thickness of the oscillating edge. The smaller the thickness, the higher the tensile stress that is applied, and thus for falsetto-like phonation one should expect higher dF/dP and for chest-like phonation where a large cross section of the vocal fold is in oscillation, a smaller dF/dP. The data in Tables II and III have frequency ranges that are similar to normal speaking phonatory ranges for humans. The canine larynges used here had vocal fold lengths of 12–16 mm and with a first-order approximation, one should be able to extrapolate the findings of this study to human phonation. Although there was no vocal tract loading on these larynges, the mechanism of pitch control via pressure may help up us to better understand the passive underlying mechanisms of pitch control in human phonation. The pressure-flow sweeps here have some resemblance to the crescendo in phonation where change in muscle contraction is not used to maintain constant pitch. The rate of pitch increase with pressure may vary in a similar fashion.
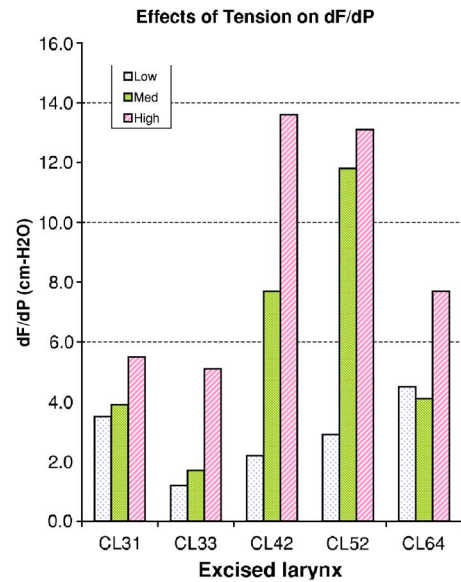
Figure 11 is a comparative chart from the dF/dP and tension data of five larynges (Table III), indicating a strong dependence of dF/dP and vocal fold tension. Two of the larynges (CL42 and CL52) showed high frequencies at medium to high tensions. Since higher tensions correspond to the higher frequencies, one can expect a good correlation between dF/dP and higher frequencies. Due to the dependence of dF/dP on both pressure and tension, a graph similar to Fig. 10 did not reveal any correlation or pattern. The tension dependence of pitch shown in Figs. 7 and 8 indicates that the excised canine model is very useful in the study of pitch control and modes of phonation. A shortcoming of the canine vocal fold is the lack of the vocal ligament, which limits the upper pitch to about 500 Hz. The effect of vocal tract loading is expected to be more on the threshold pressure than the actual range of frequency. However, this needs further investigation.

There was no single shape of the dF/dP versus F_0 data that describes the frequency rate function, but a second-order polynomial would describe these rate functions as the "V" or the "inverted V" shape and a multisegment linear model would describe them with multiple maxima and minima.

V. CONCLUSIONS

The pressure-frequency relation was obtained through a series of pressure-flow sweep experiments that were conducted using eight excised canine larynges. Vocal fold adduction and elongation were the control parameters. It was found that the pressure-frequency relation is a function whose nonlinearity is highly influenced by the level of adduction and elongation. The results indicate that dF/dP values of this study ranged from approximately 1 to 20 Hz/cm H_2O. However, for the first phonation mode, the rate of change of frequency with pressure had an average of 2.9 ± 0.7 Hz/cm H_2O, whereas for the second (higher pitch) mode the average increased to 5.3 ± 0.5 Hz/cm H_2O for adduction changes and 8.2 ± 4.4 Hz/cm H_2O for elongation

changes. The results suggest that control of F_0 as a function of P_s should take into consideration the complexity of the passive dF/dP versus F_0 relation, and that we therefore may use these relations for effective F_0 control in speech and singing.

ACKNOWLEDGMENTS

Alipour, F., Montequin, D., and Tayama, N. (**2001**). "Aerodynamic profiles of a hemilarynx with vocal tract," Ann. Otol. Rhinol. Laryngol. **110**, 550–555.

Alipour, F., Scherer, R. C., and Finnegan, E. (**1997**). "Pressure-flow relationships during phonation as a function of adduction," J. Voice **11**, 187–194.

Atkinson, J. E. (**1978**). "Correlation analysis of the physiological factors controlling fundamental voice frequency," J. Acoust. Soc. Am. **63**, 211–222.

Baer, T. (**1979**). "Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes," J. Acoust. Soc. Am. **65**, 1271–1275.

Collier, R. (**1975**). "Physiological correlates of intonation patterns," J. Acoust. Soc. Am. **58**, 249–255.

Hsiao, T. Y., Liu, C. M., Luschei, E. S., and Titze, I. R. (**2001**). "The effect of cricothyroid muscle action on the relation between subglottal pressure and fundamental frequency in an *in vivo* canine model," J. Voice **15**, 187–193.

Kataoka, K., and Kitajima, K. (**2001**). "Effects of length and depth of vibration of the vocal folds on the relationship between transglottal pressure and fundamental frequency of phonation in canine larynges," Ann. Otol. Rhinol. Laryngol. **110**, 556–561.

Kataoka, K., Kitajima, K., and Owaki, S. (**2001**). "Effects of transglottal pressure on fundamental frequency of phonation: Study with a rubber model," Ann. Otol. Rhinol. Laryngol. **110**, 56–62.

Kori, S., Sugito, M., Hirose, H., and Niimi, S. (**1990**). "Participation of the sternohyoid muscle in pitch lowering: Evidence from Osaka Japanese," Ann. Bull. Research Institute of Logopedics and Phoniatrics **24**, 65–75.

Koyama, T., Harvey, J. E., and Ogura, J. H. (**1971**). "Mechanics of voice production. II. Regulation of pitch," Laryngoscope, **81**, 45–65.

Ladefoged, P., and McKinney, N. P. (**1963**). "Loudness, sound pressure, and subglottal pressure in speech," J. Acoust. Soc. Am. **35**, 454–460.

Lieberman, P., Knudson, R., and Mead, J. (**1969**). "Determination of the rate of change of fundamental frequency with respect to subglottal air pressure during sustained phonation," J. Acoust. Soc. Am. **45**, 1537–1543.

Ohala, J. J. (**1978**). "Production of tone," *Tone, a Linguistic Survey*, edited by V. Fromkin (Academic, New York), pp. 5–39.

Rothenberg, M., and Mahshie, J. (**1986**). "Induced transglottal pressure variations during voicing," J. Phonetics **14**, 365–371.

Strik, H., and Boves, L. (**1995**). "Downtrend in F_0 and P_{sb}," J. Phonetics **23**, 203–220.

Tanaka, K., Kitajima, K., and Kataoka, H. (**1997**). "Effects of transglottal pressure change on fundamental frequency of phonation: Preliminary evaluation of the effect of intraoral pressure change," Folia Phoniatr Logop **49**(6), 300–307.

Titze, I. R. (**1989**). "On the relation between subglottal pressure and fundamental frequency in phonation," J. Acoust. Soc. Am. **85**, 901–906.

Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception

Virgilio M. Villacorta[a]
Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-591, 50 Vassar Street, Cambridge, Massachusetts 02139

Joseph S. Perkell[b]
Speech Communication Group, Research Laboratory of Electronics, and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 50 Vassar Street, Cambridge, Massachusetts 02139; and Department of Cognitive and Neural Systems, Boston University, Boston, Massachusetts 02215

Frank H. Guenther
Department of Cognitive and Neural Systems, Boston University, Boston, Massachusetts 02215 and Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-591, 50 Vassar Street, Cambridge, Massachusetts 02139

The role of auditory feedback in speech motor control was explored in three related experiments. Experiment 1 investigated auditory sensorimotor adaptation: the process by which speakers alter their speech production to compensate for perturbations of auditory feedback. When the first formant frequency (F1) was shifted in the feedback heard by subjects as they produced vowels in consonant-vowel-consonant (CVC) words, the subjects' vowels demonstrated compensatory formant shifts that were maintained when auditory feedback was subsequently masked by noise—evidence of adaptation. Experiment 2 investigated auditory discrimination of synthetic vowel stimuli differing in F1 frequency, using the same subjects. Those with more acute F1 discrimination had compensated more to F1 perturbation. Experiment 3 consisted of simulations with the directions into velocities of articulators model of speech motor planning, which showed that the model can account for key aspects of compensation. In the model, movement goals for vowels are regions in auditory space; perturbation of auditory feedback invokes auditory feedback control mechanisms that correct for the perturbation, which in turn causes updating of feedforward commands to incorporate these corrections. The relation between speaker acuity and amount of compensation to auditory perturbation is mediated by the size of speakers' auditory goal regions, with more acute speakers having smaller goal regions. © 2007 Acoustical Society of America.
[DOI: 10.1121/1.2773966]

I. INTRODUCTION

The purpose of this study is to investigate the role of sensory feedback in the motor planning of speech. Specifically, it focuses on speech sensorimotor adaptation (SA), which is an alteration of the performance of a motor task that results from the modification of sensory feedback. Such alterations can consist of "compensation"—a response to a feedback perturbation that is in the direction opposite to the perturbation, and additionally, "adaptation"—compensatory responses that persist when feedback is blocked (e.g., by masking of auditory feedback with noise) or when the perturbation is removed.

Psychophysical experiments that present human subjects with altered sensory environments have provided insight about the relationship of sensory feedback to motor control in both nonspeech and speech contexts. Experiments on limb movements have demonstrated the influence of proprioceptive feedback, i.e., feedback pertaining to limb orientation and position (Blakemore *et al.*, 1998; Bhushan and Shadmehr, 1999) and visual feedback (Welch, 1978; Bedford, 1989; Wolpert, Ghahramani and Jordan, 1995). Feedback-modification studies have also been conducted on speech production, including a number of studies that have induced compensation by altering the configuration of the vocal tract in some way (Lindblom *et al.*, 1979; Abbs and Gracco, 1984; Savariaux *et al.*, 1995; Tourville *et al.*, 2004). Other experiments have demonstrated speech compensation to novel acoustic feedback, such as delayed auditory feedback (Yates, 1963) or changes in loudness (Lane and Tranel, 1971).

Shifts of the fundamental frequency (F0) of sustained vowels have been shown to cause compensatory responses, that is, F0 modification by the speaker in the direction opposite to the shift (Kawahara, 1993; Burnett *et al.*, 1998; Jones and Munhall, 2000). Compensation for F0 shifts was especially evident when introduced during the production of tonal sequences by speakers of a tonal language (Xu *et al.*, 2004). Still others have demonstrated sensorimotor adapta-

---

[a] Current address: Irvine Sensors Corporation, Costa Mesa, CA 92626.
[b] Author to whom correspondence should be addressed. Electronic mail: perkell@speech.mit.edu

tion when vowel formants were perturbed in speakers' auditory feedback in nearly real time. For example, Houde and Jordan (1998, 2002) perturbed F1 and F2 of whispered productions of the vowel /ɛ/ along the /i/−/ɑ/ axis and found compensation that persisted in the presence of masking noise (adaptation) and generalized to other vowels. Max, Wallace and Vincent (2003) shifted all vowel formants in the same direction and showed compensation that increased with larger amounts of perturbation. Purcell and Munhall (2006) demonstrated compensation and adaptation to perturbation of F1 and F2 of voiced vowel formants. They also tracked the period following the removal of the perturbation and showed that the return to base line formant values was gradual (a "wash-out" of adaptation) and was not dependent on the number of trials during which maximal perturbation was maintained.

While introducing a vowel formant perturbation that was similar to the aforementioned paradigms, the current study builds on those earlier ones in a number of ways. The study described here: (1) utilized voiced speech (allowing for the measurement of possible fundamental frequency changes), (2) utilized a subject-dependent formant perturbation that allowed for inter-subject comparison of the degree of adaptation, (3) included female as well as male subjects, (4) measured how subjects' adaptive responses evolved over time (time-course analysis), (5) investigated the possibility of correlations between perceptual acuity and degree of adaptation, and (6) conducted simulations using a neurocomputational model of speech production that could account quantitatively for the amount and time course of compensation and adaptation. Purcell and Munhall (2006) reported results using approaches 1–4, but they did not explore the relation of compensation to auditory acuity or attempt to characterize the results with a neurocomputational model. Shifting all vowel formants in the same direction (either up or down for each subject—Max et al., 2003) essentially amounts to changing the perceived length of the vocal tract (e.g., shifting the formants up corresponds to shortening the vocal tract); whereas shifting a single formant can induce the percept of a more complex change in vowel articulation (by causing the produced vowel to sound like another vowel—also see Houde and Jordan, 1998, 2002; Purcell and Munhall, 2006).

The aforementioned evidence showing specific compensatory adjustments of speech parameters in response to perturbations of sensory feedback indicates that speech movements can make use of feedback control mechanisms. A neurocomputational model of the motor planning of speech that can be used to explore these effects is the DIVA[1] model (Guenther et al., 1998; Guenther et al., 2006). This model postulates that speech movements are planned by combining feedforward control with feedback control (cf. Kawato and Gomi, 1992) in somatosensory and auditory dimensions. The model has been shown to account for numerous properties of speech production, including aspects of speech acquisition, speaking rate effects and coarticulation (Guenther, 1995); adaptation to developmental changes in the articulatory system (Callan et al., 2000); and motor equivalence in the production of American English /r/ (Nieto-Castanon et al., 2005).

According to the DIVA model, during the initial period of speech acquisition, feedforward mechanisms are not yet fully developed, so feedback control plays a large role in ongoing speech. Through training, the feedforward controller gradually improves in its ability to generate appropriate movement commands for each speech sound (phoneme or syllable); eventually, it is the dominant controller in fluent adult speech. For mature speakers, the feedback controller is always operating, but it only contributes to motor commands when sensory feedback differs from sensory expectations, e.g. in the presence of perturbations such as the auditory modification of vowel formants introduced in the current study. If such a perturbation is applied repeatedly, the model predicts that feedforward commands will be re-tuned to account for the perturbation, and that abrupt removal of the perturbation will lead to a transient after effect (evidence of adaptation) in which the speaker still shows signs of this compensation even though the perturbation is no longer present. The DIVA model also predicts that auditory perception affects motor development such that speakers with better auditory acuity will have better tuned speech production; e.g., they will produce better contrasts between sounds. Consistent with this prediction, positive correlations between auditory acuity and produced contrast in speech have been observed for both vowels and consonants (Newman, 2003; Perkell et al., 2004a; Perkell et al., 2004b). The model further predicts that subjects with more acute auditory perception should be able to better adapt their speech to perceived auditory errors such as those introduced by F1 perturbation. The current study addresses several of these predictions.

The study comprised three experiments. The first experiment investigated auditory sensorimotor compensation and adaptation by perturbing the first formant frequency (F1) in the feedback heard by subjects as they produced vowels in CVC words. The experimental paradigm allowed us to study the time course of formant changes throughout an experimental run in vowels produced with and without masking noise. The second experiment investigated auditory acuity, measured as discrimination of synthetic vowel stimuli differing in F1 frequency, using the same subjects; this experiment was designed to determine if individuals with more acute discrimination of vowel formants also showed greater compensation to perturbations in those formants of the first experiment. The third experiment used subject-specific versions of the DIVA model of speech motor planning to simulate the subjects' performance in the first and second experiments; it was designed to determine whether the model could account quantitatively for key aspects of sensorimotor adaptation.

II. EXPERIMENT 1

This experiment was designed to test the hypothesis that human subjects utilize auditory goals in the motor planning of speech, and should modify their vowel production to compensate for acoustic perturbations in their auditory feedback. The experiment also tested the prediction that there will be adaptation: compensation that persists in the presence of masking noise and a transient after effect in which speakers

continue to show compensation for a number of trials after the perturbation is abruptly removed.

A. Real-time formant shift in vowels

A digital signal processing (DSP) algorithm was developed for shifting the first formant frequency using a Texas Instruments (TI) C6701 Evaluation Module DSP board. The algorithm utilized linear prediction coding (LPC) analysis (Markel and Gray, 1976) and a Hessenberg QR root-finding iterative algorithm (Press *et al.*, 2002) to detect the first formant (F1) in vowels. It then utilized a direct-form transpose II filter to remove the original F1, and introduced the shifted F1. This algorithm is discussed in greater detail in Appendix I and Villacorta (2006). The overall delay introduced by the digital signal processing was 18 ms, less than the 30 ms delay at which speakers notice and are disturbed by delayed feedback (Yates, 1963).

To simplify discussion of the formant shift made by the DSP board, a unit of formant shift—*perts*—is introduced here. *Perts* simply represents a multiplier of the original formant. A formant shift of 1.3 perts increased the formant to 130% of its original value (shift up), while a 0.7 perts shift decreased the formant to 70% of its original value (shift down). A pert value of 1.0 indicates that the formant was not shifted.

B. Protocol for an experimental run

The experimental run for each subject consisted of an initial calibration phase, followed by a four-phase adaptation protocol. The purpose of the calibration phase (typically 36–54 tokens in duration) was to acclimate each subject to using visual cues (target ranges and moving displays) of loudness and duration for achieving values that were needed for successful operation of the algorithm. To help assure that the subject did not hear airborne sound, insert headphones were used (see below) and the target output sound level was set at 69 dB sound pressure level (SPL) (±2 dB), significantly less than the feedback sound level of 87 dB SPL. The target vowel duration was set at 300 ms, although the actual duration could be longer due to a reaction time delay. In this phase, subjects were also questioned about the level of masking noise (87 dB SPL); as had been found in preliminary informal testing, it was determined that the level was tolerable for them and successfully prevented them from discerning their own vowel quality.

The adaptation protocol for each presentation of a token was as follows (see Fig. 1). A monitor (1 in Fig. 1) in front of the subject displayed the token (a CVC word, such as "bet") for two seconds, and also displayed the visual cues for achieving target loudness and duration. The subject spoke into a Sony ECM-672 directional microphone placed six inches from the lips (2). The speech signal transduced by the microphone was digitized and recorded for postexperiment analysis (3). The same speech signal was sent concurrently to the TI DSP board for the synthesis of formant-shifted speech (4). The output of the DSP board (formant-shifted speech) was sent to a feedback selector switch which determined, depending on which token was presented to the subject,
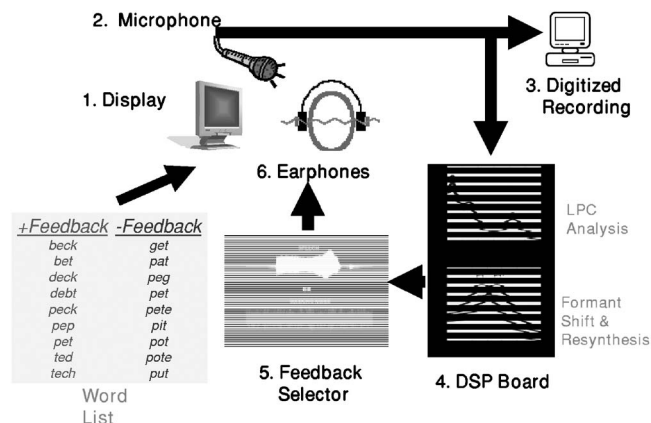
FIG. 1. Schematic diagram of the cycle that occurred during the presentation of one token during an SA experimental run. Refer to Sec. II B for a detailed description.

whether the subject heard masking noise or the perturbed speech signal (5). The appropriate signal was then presented to the subject over EarTone 3A insert earphones (Ear Auditory Systems) (6)[2]. The perturbed speech signal from the DSP board and the output signal from the selector switch were also digitized and saved for postexperimental analysis.

A total of 18 different target words ("Word List" in Fig. 1 and Table I) were used. The experiment consisted of a number of *epochs*, where each epoch contained a single repetition of each of the 18 target words. Nine of these words (+*feedback*) were presented with the subjects able to hear auditory feedback (either perturbed or unperturbed, depending on the phase of the experiment) over the earphones; all of these words contained the vowel /ɛ/ (the only vowel trained). The other nine words (−*feedback*) were presented with masking noise. Three of the −*feedback* words contained the vowel /ɛ/, one in the same phonetic context as the word presented in the +*feedback* list ("pet") and two in different phonetic contexts ("get" and "peg"). The other six −*feedback* words contained vowels different from the training vowel. The order of the +*feedback* tokens and −*feedback* tokens was randomized from epoch to epoch; however, all of the +feedback tokens were always presented before the −*feedback* tokens within an epoch.

For each subject, the adaptation protocol comprised four phases: *base line*, *ramp*, *full perturbation* and *postperturbation* (schematized in Fig. 2). Each phase consisted of a fixed number of epochs. The base line phase consisted of the first 15 epochs, and was performed with the feedback set at

TABLE I. Word list for the SA experiment.

+Feedback	−Feedback
beck	get
bet	pat
deck	peg
debt	pet
peck	pete
pep	pit
pet	pot
ted	pote
tech	put

2308   J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Villacorta *et al.*: Sensorimotor adaptation and perception of vowels

FIG. 2. Diagram of the level of F1 perturbation presented during one experimental session, as a function of epoch number (where an epoch consists of one repetition of each of the 18 words in the corpus). The 65 epochs of an experimental session are divided into four phases (demarcated by dashed vertical lines). From left to right, these phases are *base line* (epochs 1–15), *ramp* (epochs 16–20), *full perturbation* (epochs 21–45), and *postperturbation* (epochs 46–65). The protocols for two subject groups are shown: those undergoing an upward F1 shift (upper line) and those undergoing a downward F1 shift (lower line).

1.0 pert (no formant shift). The following ramp phase (epochs 16–20) was used to gradually introduce the formant shift by changing the pert level by 0.05 pert per epoch. Depending on the subject group (*shift up* or *shift down*—see below), during the full perturbation phase (epochs 21–45), the speech feedback had either a 1.3 pert shift or a 0.7 pert shift. During the entire postperturbation phase (epochs 46–65), the feedback was again set at 1.0 pert (no shift); this phase allowed for the measurement of the persistence of any adaptation learned during the full-perturbation phase. An entire experiment for one subject consisted of 65 epochs, comprising a total of 1170 tokens; the experiment lasted approximately 90–120 min.

C. Subject selection criteria and description

Subjects were 20 adult native speakers of North American English with no reported impairment of hearing or speech. Five females and five males were run with an upward F1 shift (*shift-up* subjects); another five females and five males were run with a downward F1 shift (*shift-down* subjects). The subjects had an age range from 18 to 44 with a median age of 21. Informed consent was obtained from all subjects.

D. Postexperiment spectral analysis of tokens

Following the experiment, a spectral analysis was performed on the speech signals that had been digitized directly from the microphone. Each recorded token (sampled at 16 kHz) was labeled manually at the beginning and end of the vowel on the sound-pressure wave form; then the first two formants were extracted utilizing an automated algorithm designed to minimize the occurrence of missing or spurious values. Formants were derived from an LPC spectrum taken over a sliding 30 ms window. The spectrum was measured repeatedly between 10% and 90% of the delimited vowel interval in 5% increments, and the mean formant values 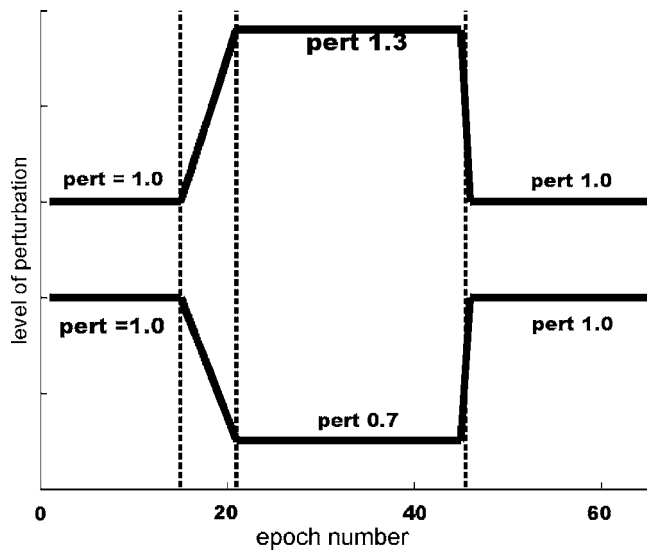over these repeated measures were recorded. The analysis for a majority of the subjects used an "optimal" LPC order determined by a heuristic method that utilizes a reflection coefficient cutoff (Vallabha and Tuller, 2002). For subjects with a large number of missing or spurious formants, the analysis was repeated using LPC orders of 14–17 inclusive.

The fundamental frequency (F0) was extracted from each token using a pitch estimator that is based on a modified autocorrelation analysis (Markel *et al.*, 1976). For some tokens, F0 appeared to be underestimated, so F0 values below 50 Hz were excluded from analysis. For all but one subject, this exclusion criterion removed less than 3% of the tokens. One subject had 44% of tokens excluded by this criterion, so that subject's data were excluded from the F0 analysis.

To allow comparison among subjects with differing base line formant frequencies and F0, especially differences related to gender, each subject's formant and F0 values were normalized to his or her mean base line values, as shown in Eq. (1) for F1.

$$\text{norm\_}F1 = \frac{F1_{\text{Hertz}}}{\text{mean}(F1)_{\text{base line phase}}}. \tag{1}$$

In order to compare changes from the base line (normalized value =1.0) to the full-pert phase among all the subjects (regardless of the direction of the F1 shift), an *adaptive response index* (ARI) was calculated as shown in Eq. (2). Larger, positive ARI values indicated greater extent of adaptation for that subject, while negative ARI values (which occurred for two of the 20 subjects) indicated that those subjects produced responses that followed the perturbation, rather than compensated for it.

$$\text{ARI} = \begin{cases} \text{mean(norm\_}F1 - 1)_{\text{full pert phase}}, & \text{if } pert = 0.7 \\ \text{mean}(1\_\text{norm} - F1)_{\text{full pert phase}}, & \text{if } pert = 1.3 \end{cases}. \tag{2}$$

E. Results

Figure 3 shows normalized F1 (solid curves) and F2 (dashed curves) values for the *+feedback* tokens averaged across all subjects in each group.[3] Data from shift-down subjects are shown with black lines; from shift-up subjects, with gray lines. The error bars show one standard error about the mean. The figure shows that subjects compensated partially for the acoustic perturbation to which they were exposed. Shift-up subjects increased vowel F1 during the experiment (black solid line), while shift-down subjects decreased F1 (gray solid line).[4] Compared to the changes in F1, F2 changed by very small amounts.

Generally, subjects responded with only a short delay to the acoustic perturbation: the first significant change in normalized F1 occurred during the second epoch in the ramp phase (epoch 17). This compensation was retained for some time after the perturbation was turned off at epoch 45 (i.e., during the *postpert* phase), indicating that subjects had
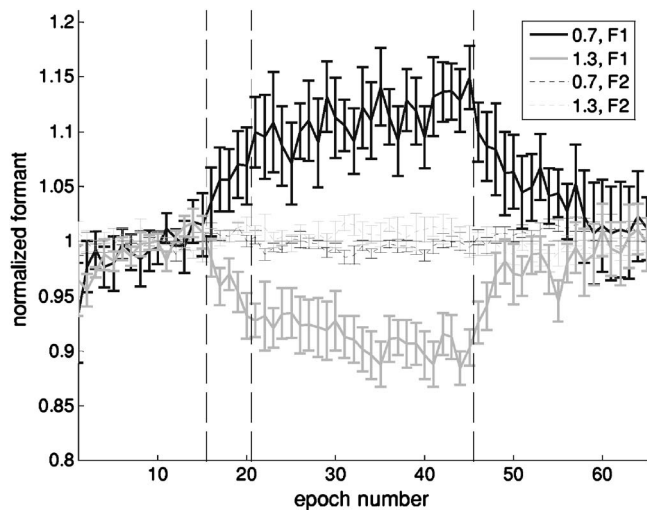
FIG. 3. Produced first and second formant frequencies, normalized to the adjusted base line, as a function of epoch number in +*feedback* words for all subjects. The upper curve corresponds to the normalized F1 for the ten subjects run on the shift-down protocol; the lower curve corresponds to the shift-up protocol. Each data point is the mean value of the nine +*feedback* words across ten subjects (five male, five female). The dashed vertical lines demarcate the phases of the protocol; the dashed horizontal line corresponds to base line values. Normalized F2 values are shown as the dashed curves, which remain close to the base line value of 1.0. The error bars depict the standard error of the mean among ten subjects.

adapted to the perturbation. Normalized F1 consistently returned to base line within the standard error after epoch 55, approximately 15–20 min into the postpert phase. This finding is consistent with those of Purcell and Munhall (2006), who also showed that recovery to base line formant frequencies was not immediate when the formant perturbation was removed.

The extent of adaptation was less than the amount required to fully compensate for the acoustic perturbation. For shift-down subjects, full compensation (i.e., the inverse of 0.7) would be represented by a normalized F1 value of 1.429; the greatest actual change for the shift-down subjects had a mean normalized value of 1.149 (i.e., approximately 35% compensation), which occurred in epoch 45. Similarly, full compensation for the shift-up subjects (1.3 pert shift) would be represented by a normalized F1 value of 0.769. Their greatest change had a mean normalized value of 0.884 (approximately 50% compensation), which occurred in epoch 44.

The –*feedback* tokens were analyzed in the same way to determine the extent to which adaptation would occur for the same vowel with auditory feedback masked (that is, without perception of the perturbed signal). As mentioned above, the word list contained tokens that were uttered with auditory feedback masked, but which contained the same vowel the subjects had heard with full perturbation (/ɛ/). The DIVA model predicts that adaptation learned for /ɛ/ with feedback perturbed should be maintained even without acoustic feedback. Indeed, in their SA study of with whispered vowels, Houde and Jordan (1998, 2002) demonstrated that such adaptation was maintained in the absence of acoustic feedback and also that it generalized to productions of the same vowel in different phonetic contexts. The current –*feedback* adap-
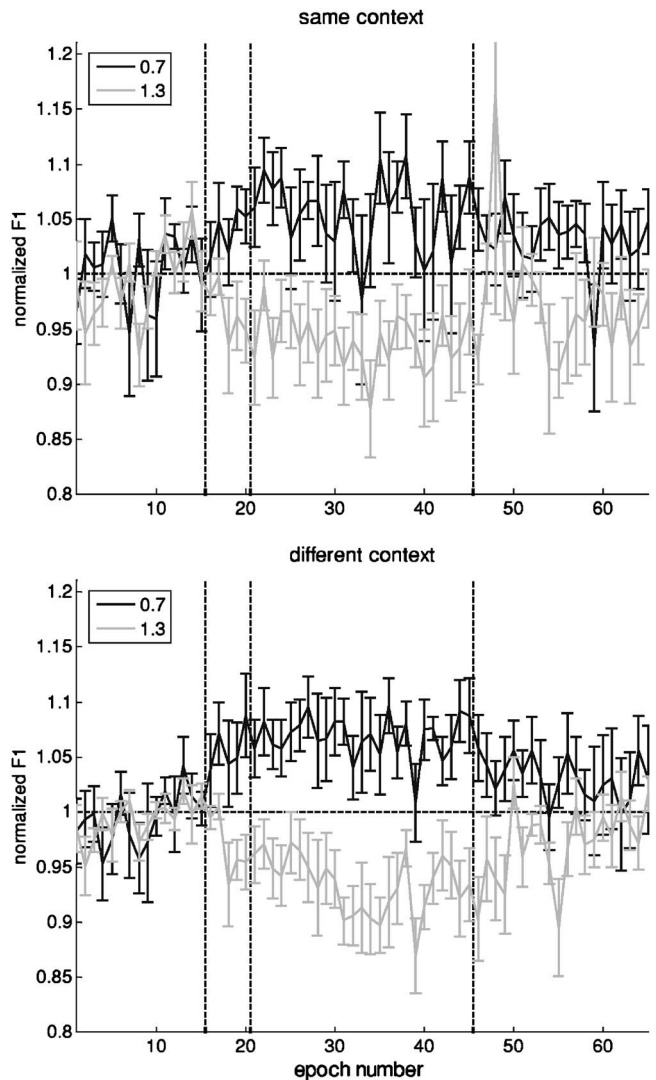
FIG. 4. Produced first formant frequency, normalized to the base line, in the –*feedback* words containing the vowel /ɛ/. The top plot shows normalized F1 for the *same context*, –*feedback* token ("pet"), while the bottom figure shows normalized F1 for the *different context*, –*feedback* tokens ("get" and "peg"). The axes, data labels and vertical markers are the same as in Fig. 3, except that normalized F2 is not shown.

tation results for /ɛ/ are divided into two groups: –*feedback* adaptation for the *same context* token, and –*feedback* adaptation for *different context* tokens. The *same context* token—referring to the fact that this token is also contained in the +*feedback* word list—is the token "pet." The *different context* tokens are the tokens "get" and "peg," which were not present in the +*feedback*, word list.

Figure 4 shows that the adaptation to perturbation of +*feedback* /ɛ/ tokens does indeed occur for the *same context*, –*feedback* token. However, adaptation in the –*feedback* tokens occurred to a lesser extent than in the +*feedback* tokens (compare with Fig. 3). This finding is confirmed by comparing ARI values (Eq. (2)) between +*feedback* tokens and –*feedback* tokens. The *ARI* for the –*feedback, same context* condition was 58% of the ARI for the +*feedback* tokens, which is a significant difference ($t[198]=2.3, p<0.05$). Additionally, the ARI in the –*feedback, different context* condition was 67% of the +*feedback* (ARI) condition, which is also a significant difference ($t[218]=2.47, p<0.05$). While
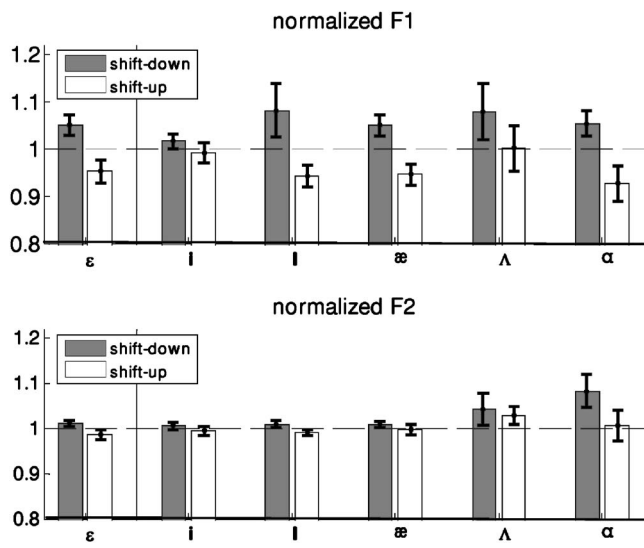
FIG. 5. Full-pert phase formants normalized to base line for all −*feedback* token vowels. Mean first formant values are shown in the upper plot; second formant values, in the lower plot. Values from shift-down subjects are represented by dark bars and from shift-up subjects, by light bars. Error bars show standard error about the mean.

FIG. 6. Normalized F0 as a function of epoch number. To maintain consistency with Fig. 3, only +*feedback* vowels are shown. The solid line represents the mean values from the shift-down subjects; the dashed line represents the mean values from the shift-up subjects. The vertical lines demarcate the phases of the experiment.

the mean changes in F1 for the *different context* tokens appear to be greater than the F1 changes for the *same context* tokens, there were no significant differences between the two context conditions for both the shift-down and shift-up subjects.

Several −*feedback* tokens contained vowels different from the one subjects produced with feedback perturbed (/ɛ/). These tokens were included in the protocol to establish the degree to which adaptation would generalize to unperturbed vowels. The bar plots in Fig. 5 display the amount of adaptation found for the following vowels: /ɪ/ ("pit"), /i/ ("pete"), /æ/ ("pat"), /ʌ/ ("put"), and /ɑ/ ("pot").[5] The −*feedback* token /ɛ/ is also displayed for comparison. Shown are the mean F1 (upper plot) and F2 (lower plot) of these vowels, normalized with respect to each vowel's base line formant values.

For most vowels, the mean normalized F1 was significantly above the base line in shift-down subjects, and was significantly below the base line in shift-up subjects ($p < 0.01$). However, the vowels /i/ and /ʌ/ did not show consistent F1 generalization. The shift-down subjects demonstrated vowel a small significant increase of F1 for /i/ ($t[249]=2.33, p < 0.05$); the shift-up subjects showed a small decrease in F1 for /i/ that was not significant. For the vowel /ʌ/, the shift-down subjects demonstrated a significant upward F1 shift, but the shift-up subjects failed to demonstrate a significant decrease. (Villacorta, 2006, shows that this lack of generalization for /ʌ/ is due to the male, shift-up subgroup.) As seen in the lower plot, changes in F2 were considerably smaller in magnitude than in F1 ($p < 0.05$), demonstrating formant specificity of the generalization for most of the vowels. The vowel /i/ did not show a significant F2 change for either shift-down or shift-up subjects—likely due to the fact that the F1 changes for /i/ were also relatively small. The vowel /ɑ/ did not show significantly smaller F2 changes (compared to F1 changes) in the shift-down sub-

jects. Anomalously, the vowel /ʌ/ showed F2 increases for the shift-down as well as the shift-up subjects (possibly related to the above-mentioned outlying F1 responses of the male, shift-up subgroup).
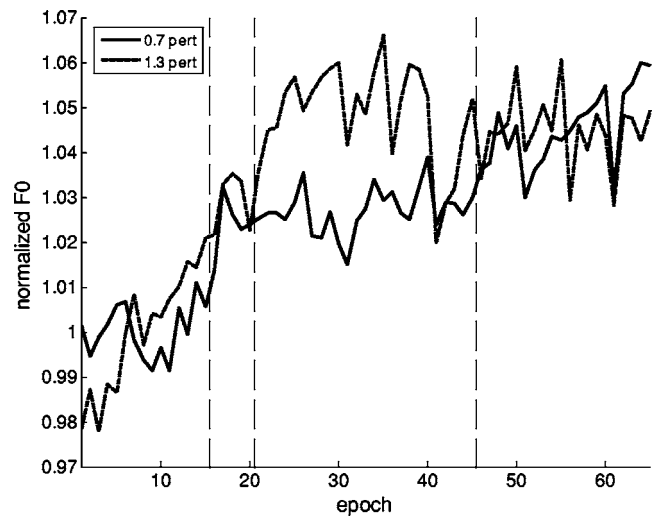
Figure 6 shows F0 as a function of epoch number, averaged across shift-up (dashed line) and across shift-down subjects (solid line) and normalized to the mean of the base line epochs. The figure shows that both shift-down and shift-up subjects demonstrated a general trend of increasing F0 throughout the experiment. The relation between changes in F0 and F1 (factoring out the common upward trend in F0) was investigated by calculating the difference between subject F0 value and the mean F0 and the difference between subject F1 and mean F1 across all subjects at each epoch. It was found that subjects modified F0 in a direction opposite to the compensatory F1 shift they produced; this relation was highly significant ($r=-0.74, p < 0.001$). It is possible that the duration of each utterance (300 ms), the large number of utterances produced by each subject (approximately 1170 tokens), and the overall duration of the experiment (90–120 min) all combined to cause fatigue that led to an upward drift in F0. Some support for this claim can be inferred from a similar upward F0 drift observed by Jones and Munhall (2000).[6]

Analysis of the adaptive response index values for F1 and F2 showed that, from the ramp phase through the post-pert phase, the direction of the small AR_{F2} change appears to be opposite to AR_{F1} changes. The mean AR values across all subjects for this subset of epochs (ramp phase through post-pert phase), showed a significant inverse relation between AR_{F1} and AR_{F2} ($r=-0.78, p < 0.001$). Thus the observed changes in F0, F1 and F2 lead to the inference that the auditory space in which subjects adapt is characterized by dimensions that depend on multiple formants and F0.

III. EXPERIMENT 2

To investigate whether subjects' auditory acuity was related to the amount of their adaptation, a second experiment

was conducted to measure auditory acuity of F1 variation with the same subjects who served in Experiment 1. This experiment consisted of three parts: (1) a recording of the subject's "base" tokens, (2) an adaptive staircase discrimination task and (3) a second, more finely tuned discrimination task. The end result was a measure of each subject's auditory acuity. The use of a two-stage protocol for obtaining an accurate estimate of auditory acuity was based on prior work (Guenther *et al.*, 1999a; Guenther *et al.*, 2004).[7]

A. Participating subjects

The subjects were a subset of those who participated in Experiment 1. Seven out of the original 20 subjects were no longer available at the time Experiment 2 was conducted, so the results from the acuity experiment were based on the 13 subjects who could be recalled. Informed consent for the auditory acuity experiments was obtained from all subjects.

B. Recording of the subject's speech

Subject-specific synthetic stimuli were used for the acuity tests. For this purpose, each subject was recorded while speaking ten tokens each of the words "bet," "bit" and "bat." The recordings were conducted in a sound attenuating room using a head-mounted piezo-electric microphone (Audio-Technica, model AT803B) placed at a fixed distance of 20 cm from the speaker's lips. Elicited utterances were presented on a monitor. As in Experiment 1, the monitor also displayed cues that induced the subject to speak at a target loudness (85±2 dB SPL) and word duration (300 ms). Subjects were allowed to practice to achieve these targets. The F1 frequency for each "bet" token was measured, and the "bet" token with the median F1 value was used to determine the F1 of a *base token*. Synthetic vowels varying in F1 were generated offline using a MATLAB program that ran a formant perturbation algorithm identical to what was run on the TI DSP board.

The acuity tests were carried out in the same sound attenuating room in which the recordings were made, though not always on the same day. Subjects heard stimuli over closed-back headphones (Sennheiser EH2200), played on a computer controlled by a MATLAB script.

C. Staircase protocol for estimation of *jnd*

In an initial stage of acuity testing, a staircase protocol was used to rapidly obtain an approximate estimate of the just noticeable difference (*jnd*) in F1 for each subject. This estimate was then used to determine a narrower range of tokens for the second stage, which utilized a larger number of trials with token pairs that were chosen to fall near the subject's initial *jnd*, in order to produce a more accurate estimate of auditory acuity.

An adaptive, one-up, two-down staircase protocol was run to estimate the *jnd* for F1 around the *base token* obtained from the subject's speech recording (as illustrated in Fig. 7). In this procedure, pairs of tokens that were either the *same* or *different* from each other were presented to the subject with equal probability. The *same* pairs consisted of repetitions of the *base token*, while the *different* pairs consisted of tokens
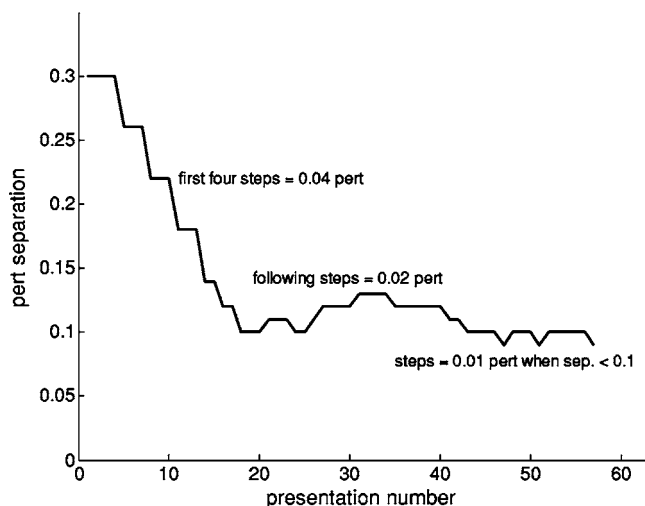
FIG. 7. Example of the adaptive procedure used to estimate *jnd*. The abscissa shows the presentation number of the given pair, and the ordinate depicts the separation of the *different* pairs in pert. The text within the figure gives conditions for changes in step size. The staircase terminated after eight reversals.

with F1 values greater or lesser than that of the *base token*, equally spaced in pert. For example, the *different* pair separated by 0.3 pert consisted of the 0.85 pert and the 1.15 pert tokens. Whenever the subject responded incorrectly to either the *same* or *different pairs*, the distance between the members in the *different* pairs increased. Whenever the subject responded correctly to two presentations of a given different pair, the distance between the members of the different pairs decreased. The separation was unchanged when the subject responded correctly to a *same*, pair presentation.[8]

D. Determining auditory acuity

A more precise protocol involving many more *same-different* judgments was then run on each subject. In the *jnd* protocol, presented tokens were either the *same* (with both tokens equal to the base token) or *different* (straddling the base token). The *different* pairs were spaced by the following multiples of the jnd_{est}: ±0.25, ±0.5, ±0.75, ±1.0 and ±1.4. The +multiple of the jnd_{est} pair (e.g., +0.25, +0.5) was always presented with the corresponding multiple (e.g., −0.25, −0.5) for a *different* pair presentation, though the order of the tokens within a pair was randomized (e.g., +0.25 followed by −0.25 or −0.25 followed by +0.25). Each unique pair (the single *same* and each of the five *different* pairs) was presented to the subject 50 times, for a total of 300 presentations per block. Subjects were given feedback consisting of the correct response to the pair just presented.

A d' score for each pair was calculated using a standard signal detection theory formula (Macmillan and Creelman, 2005) shown in Eq. (3), where z is the normal inverse function, H is the hit rate (responds *different*|different) and F is the false alarm rate (responds *different*|same). Note that all rates were calculated as a fraction of a total of 50.5 presentations (rather than 50 presentations) to avoid undefined z scores.

$$d' = z(H) - z(F) \qquad (3)$$

Villacorta *et al.*: Sensorimotor adaptation and perception of vowels
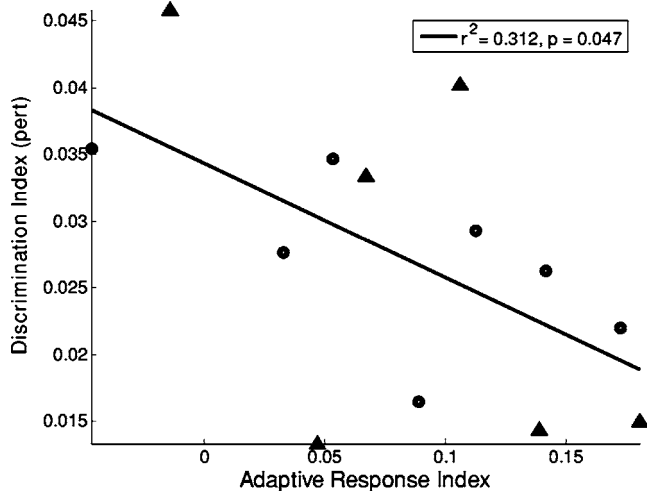
FIG. 8. The adaptive response index is correlated with the *jnd* score of the base token. The ordinate shows the *jnd* score (*Discrimination Index*), while the abscissa shows the *adaptive response index*. The open circles represent shift-down subjects, while the triangles represent shift-up subjects. Statistics for the regression line are shown in the legend.

Data consisting of d' score as a function of pair separation (in perts) were then fitted with a sigmoid function. A sigmoid function was used in this case because it is monotonic and best captures the sharp rise of d' in the sensitive region, while also capturing ceiling and floor effects observed in the data. To estimate perceptual acuity, a "discrimination index" (DI) was calculated from the sigmoid fit to the d' function. We defined the DI as the separation (in perts) that corresponds to a d' of 0.7. (A d' of 0.7 was used here because it was the maximum d' value common to all subjects run on the perceptual acuity protocol.) Note that the larger the DI, the worse the subject's acuity (i.e., the further apart two stimuli need to be for detection by the subject).

E. Results

The subjects' DIs were significantly correlated with their *adaptive response indices*, as shown in Fig. 8. This figure shows DI as a function of ARI for the shift-down subjects (open circles) and the shift-up subjects (triangles), along with a regression line. The line demonstrates the predicted trend: subjects with smaller *jnds* tend to adapt to a greater extent. The relation between *jnd* and adaptive response was significant ($r = 0.56$, $p < 0.047$), accounting for 31% of the variance.

It was observed that the produced F1 separation between neighboring vowels varied from subject to subject, which could have a confounding influence on the extent of adaptation measured during the SA experiment and therefore on the correlation with *jnd* values. Since the SA experiment included base line (epochs 1–15) tokens of the vowels /æ/, /ε/, and /ɪ/ ("pat," "pet," and "pit" used as −feedback tokens), it was possible to measure the F1 separation in neighboring vowels and subsequently control for it. Equation (4) shows how normalized vowel separation in F1 was calculated. Note that the *F1_separation* values are normalized by the base line F1 from the word "pet," and that only −*feedback* base line tokens were used for this measurement.
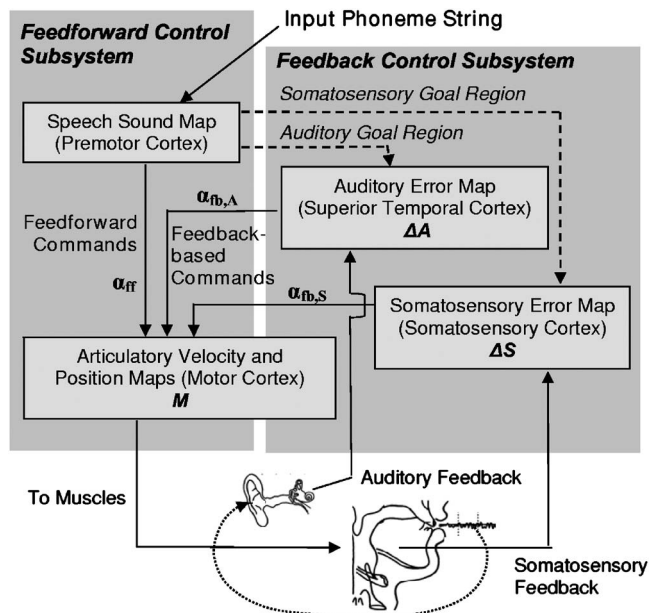
FIG. 9. A functional diagram of the DIVA model of speech motor control. The feedforward component projects from the speech sound map P, and is scaled by weight α_{ff}. The feedback component consists of projections from the auditory (ΔA) and somatosensory (ΔS) error maps, and are scaled by weights $\alpha_{fb,A}$ and $\alpha_{fb,S}$, respectively. The feedforward and feedback projections are integrated by the speech motor cortex M to yield the appropriate speech motor commands, which drive the vocal-tract articulators to generate speech sounds.

$$F1\_separation_{\text{pet–pit}} = \frac{\text{pet}\_F1_{\text{median}} - \text{pit}\_F1_{\text{median}}}{\text{pet}\_F1_{\text{median}}}$$

$$F1\_separation_{\text{pat–pet}} = \frac{\text{pat}\_F1_{\text{median}} - \text{pet}\_F1_{\text{median}}}{\text{pet}\_F1_{\text{median}}}.$$

(4)

For a given subject, the relevant $F1\_separation$ value was the one characterizing the separation between the two neighboring vowels corresponding to the direction of perturbation used in the SA protocol. Therefore, $F1\_separation_{\text{pet–pit}}$ was used for the shift-down subjects and $F1\_separation_{\text{pat–pet}}$ was used for the shift-up subjects.

The partial correlation coefficient ($r_{x,y\|z}$) represents the correlation between two measures (DI and ARI) when controlling for *normalized* $F1\_separation$. This statistic, $r_{\text{acuity\_index ARI\|norm\_F1\_separation}}$ had a highly significant value ($r = 0.79$; $p < 0.001$), accounting for over 62% of the variance and indicating that smaller *jnd* values (i.e., greater perceptual acuity) are associated with larger adaptation scores.[9]

IV. EXPERIMENT 3

This experiment was designed to compare simulations using the DIVA model of speech motor planning to the human subject results from the SA and auditory acuity studies. Figure 9 shows a simplified schematic diagram of the DIVA model, indicating the relation between feedback and feedforward control of speech movements in the cerebral cortex. The model is described here briefly; it is discussed in depth in Guenther *et al.* (2006).

The *speech sound map* (hypothesized to lie in left pre-

motor cortex) projects sensory expectations associated with the current speech sound to auditory (ΔA) and somatosensory (ΔS) error cells, where these expectations (or *goals*) are compared to the actual sensory feedback. The projections of sensory expectations are learned and improve with practice. The output from the sensory error cells projects to an articulatory velocity map, resulting in the feedback-based component of the motor command; the gains $\alpha_{\text{fb},A}$ and $\alpha_{\text{fb},S}$ control how much each feedback source contributes to the overall motor command.

The speech sound map—aside from giving rise to the sensory expectations projecting to the sensory error cells—also projects directly to motor cortex, giving rise to a feedforward component of the motor command. By incorporating the results of previous attempts to produce the given speech sound with auditory feedback available, this motor command improves over time.

The feedforward and the two feedback components of the motor command are integrated to form the overall motor command M, which determines the desired positions of the speech articulators. The motor command M in turn drives the articulators of the vocal tract, producing the speech sound; this production provides sensory feedback to the motor control system. For use in simulations, the DIVA model's motor commands, M, are sent to an articulatory based speech synthesizer (Maeda, 1990) to produce an acoustic output.

When the model is first learning to speak (corresponding to infant babbling and early word production), the feedback component of speech control plays a large role, since the model has not yet learned feedforward commands for different speech sounds. With continued speech training, the feedforward projections from the speech sound map improve in their ability to predict the correct feedforward commands. In trained fluent (e.g., adult) speech in normal conditions, feedforward control dominates the motor command signal since the error signals resulting from the auditory and somatosensory error cells are small due to accurate feedforward commands. Alterations in auditory feedback—as introduced by the SA protocol—produce mismatches between expected and actual auditory consequences, which results in an auditory error signal. This causes the feedback control signal (specifically the auditory component) to increase and significantly influence the output motor commands. Adaptation occurs in this model as the feedforward projections are adjusted to account for the acoustic perturbation.

In the SA protocol, only the auditory component of the sensory feedback is perturbed; the somatosensory feedback is left unperturbed. The model predicts that adaptation should not fully compensate for purely auditory perturbations due to the influence of somatosensory feedback control. That is, as the feedforward commands change to compensate for the auditory perturbation, somatosensory errors begin to arise and result in corrective motor commands that resist changes in the feedforward command. As observed above, analyses from the +*feedback* tokens of the SA subjects also demonstrated only partial compensation (refer to Sec. I B), supporting the model's prediction.

A. Modeling variation in auditory acuity

One important property of the DIVA model is its reliance on sensory goal *regions*, rather than *points* (Guenther *et al.*, 1998; Guenther, 1995). The notion of sensory goal regions explains a number of phenomena related to speech production. These observed behaviors include motor equivalent articulatory configurations (Guenther, 1995; Guenther *et al.*, 1998) and their use in reducing acoustic variability (Guenther *et al.*, 1998; Guenther *et al.*, 1999b; Nieto-Castanon *et al.*, 2005), as well as anticipatory coarticulation, carryover coarticulation, and effects related to speaking rate (Guenther, 1995).

Prior studies have demonstrated that speakers with greater auditory acuity produce more distinct contrasts between two phonemes (Newman, 2003; Perkell *et al.*, 2004a, b). According to the DIVA model, these larger contrasts result from the use of smaller auditory goal regions by speakers with better acuity; this may occur because these speakers are more likely to notice poor productions of a sound and thus not include them as part of the sound's target region. In keeping with this view, we created a version of the model for each individual subject by using an auditory target region size for the vowel /ε/ that was proportional to the subject's discrimination index. The details of this process are described in Appendix II. In short, subjects with a larger discrimination index (reflecting poorer acuity) were modeled by training the DIVA model with large target regions, while subjects with better acuity were modeled by training on smaller target regions. These varying trained models were then used in a simulation experiment that replicated the sensorimotor adaptation paradigm of Experiment 1.

B. Design of the SA simulations within the DIVA model

Twenty simulations were performed, using subject-specific versions of the DIVA model; each simulation corresponded to a particular subject's SA run, with the model's target region size adjusted using the relation between acuity and adaptive response described in Appendix II. Each simulation consisted of the same four phases as the human subject SMA experiment: *base line, ramp, full pert*, and *post pert*. During these phases, auditory feedback to the model was turned on and off to replicate the +*feedback* and −*feedback* SA results. Like the human subject experiment, the perturbation to F1 in the model's auditory feedback during the *full-pert* phase was either 0.7 or 1.3 pert (depending on the subject being simulated), and the perturbation was ramped up during the ramp phase as in the experiment.

In the SA experiment with human subjects, each epoch contained nine +*feedback tokens* and three −*feedback* tokens that contained the vowel /ε/. To maintain this ratio while simplifying the simulations, one epoch in the simulation was composed of four trials: three trials with feedback turned on, followed by one trial with feedback turned off.

C. Results

Figure 10 compares the results from +*feedback* trials in the DIVA simulations to the corresponding human subject
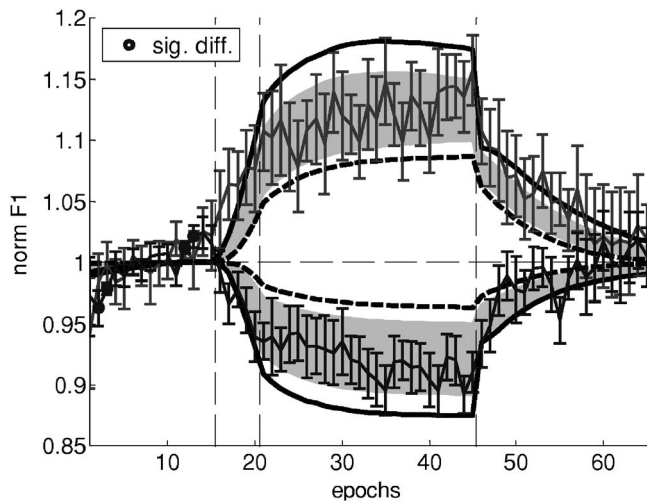
FIG. 10. Normalized F1 as a function of epoch number during the SA protocol in +*feedback* trials: DIVA simulations compared to human subject results. The thin lines shown with standard error bars correspond to the subject SA data (20 subjects). The shaded region corresponds to the DIVA simulations, and represents the 95% confidence interval about the mean. The vertical dashed lines show the experiment phase transitions; the horizontal dashed line indicates base line. The open circles indicate epochs in which the data and the simulation results were significantly different. The black solid curves correspond to high-acuity simulations, while the black dashed curves correspond to low-acuity simulations.

FIG. 11. Normalized F1 during the SA protocol in −*feedback* trials: DIVA simulations compared to subject results.

data. These results demonstrate that the SA simulations account for the main trends found in the human SA data: (1) a compensatory change in F1 that builds gradually over the *ramp* and *full pert* phases, (2) a maximum F1 deviation that only partially compensates for the perturbation, and (3) a gradual return to the base line F1 value in the *postpert* phase. Furthermore, acuity and the extent of F1 deviation are positively related in the model, evident by comparing the high acuity (solid lines) to the low acuity (dashed lines) simulations, as in the human subject data (not shown in Fig. 10). Finally, there is a slight asymmetry between the shift-up group and shift-down group, seen in both the simulations and the human subject results. This is not surprising, given that the inverse of the perturbation—which represents the maximal response expected—is a larger change from base line for the shift-down condition than for the shift up.

To determine if the simulation results were significantly different from the human subject results, a pooled, two-tail t test was performed on an epoch-by-epoch basis between the two sets of results; differences statistically significant at a level of $p=0.05$ are indicated in Fig. 10 by the open circles. The simulation results differed significantly only during four epochs, all of which were in the *base line* phase, where the experimental subjects showed considerable drift in F1 compared to the constant F1 of the model's productions. During the ramp phase, the human SA results seem to show a faster adaptive response than the simulation results, but this difference is not statistically significant.

Like the human subject results, the DIVA simulations produced very little change in the second formant (not shown): the normalized F2 during the full-pert phase had a mean value of $1.0135+/−0.0035$ for the shift-down simulations, and a mean value of $0.9975+/−0.0004$ for the shift-up simulations. It should be noted that the simulations and the
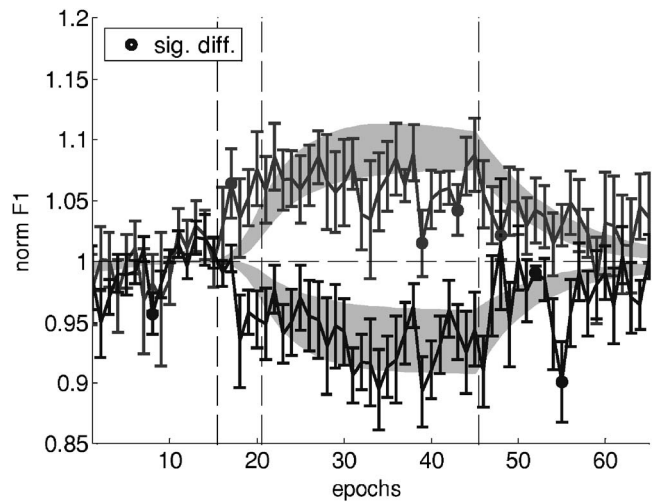
humans subject results differed in the direction of the F2 changes; unlike the human subjects, who showed F1 and F2 shifting in opposite directions, the simulations showed changes in F1 and F2 occurring in the same direction. As described earlier, the shifting of F1 and F2 in opposite directions by the experimental subjects may indicate the use of an auditory planning frame that is not strictly formant based as implemented in the model simulations, but rather is better characterized by relative values of the formants and F0.

Figure 11 compares the results from −*feedback* trials in the DIVA simulations to the corresponding human subject data. The simulations exhibit adaptive responses that are similar in extent to those seen in human data in −*feedback* tokens. Excluding differences in the base line phase, the −*feedback* simulations differed from the human subject data in four epochs for the shift-down condition (one epoch in the ramp phase, two in the full-pert phase and one in the postpert phase), and in two epochs for the shift-up condition (both in the postpert phase). It should be noted that, because corrections for multiple comparisons were not done in order to make the test of the model more stringent, one would expect 2–3 epochs out of 50 to show "false" significant differences (for a significance threshold of $p=0.05$) even if the statistical distributions of model and subject productions were identical.

V. DISCUSSION

The studies presented in this article reveal several details of the process by which individuals modify speech in order to compensate for altered acoustic feedback. The results from Experiment 1 indicate that, in response to perturbations of the first formant (F1) in the acoustic feedback of vowel productions, subjects compensate by producing vowels with F1 shifted in a direction opposite to the perturbation. Specifically, shift-down subjects exhibited 35% compensation, and shift-up subjects exhibited 50% compensation. This range of compensation is similar to other experiments in vowel formant manipulation (Houde *et al.*, 1998, 2002; Max *et al.*, 2003). Although we observed an asymmetry in compensation relative to the direction of shift, this asymmetry arises from

how we defined complete compensation (see Sec. I B). Compensation was present even when the subjects' auditory feedback was blocked by masking noise—defined as true adaptation in Houde and Jordan (2002)—and it generalized to most other vowels that were not perturbed. This adaptation persisted for a period (roughly 10 epochs) after the perturbation was removed, decaying in a similar manner to that shown in Purcell and Munhall (2006).

Previous vowel formant SA experiments also exhibited inter-subject variation in the extent of adaptation, which was found in our experiment. Unlike those aforementioned experiments, we also measured auditory acuity to formant differences in a majority of our subjects (Experiment 2), allowing us to study a possible source of this inter-subject variation. Indeed, the results of Experiment 2 show that auditory acuity and compensatory responses to perturbations of F1 are significantly correlated: subjects with greater acuity demonstrated greater responses. This correlation increased when factoring out inherent speaker differences in the separation in F1 for neighboring vowels. This finding is similar to previous evidence of a linkage between the discrimination of vowel and sibilant acoustics and the production of those sounds with greater acoustic contrast (Perkell *et al.*, 2004b; Newman, 2003; Perkell *et al.*, 2004a); in the current study, "production contrast" is reflected in greater compensatory change in the first formant frequency. The strength of the correlation between subject acuity and amount of compensation when individual vowel spacing is factored out ($r = -0.79$; $p < 0.001$) is noteworthy, especially considering possible confounds such as imperfect performance of the SA signal processing algorithm and variability inherent in the methods used for the perceptual testing.

In the simulations, the DIVA model was able to account quantitatively for several key characteristics of the human subject results. Similar to the human subject results, the DIVA simulations demonstrated formant-specific compensation to acoustic perturbation, adaptation in the absence of auditory feedback and persistence for a short period after the perturbation was removed. Such adaptation can be attributed to modification of feedforward commands that washes out once the source of auditory errors has been removed.

Individual auditory acuity for vowels is reflected in the DIVA model by auditory goal regions of varying size, which are smaller in individuals with greater acuity. The simulations also demonstrated that smaller auditory goal regions—hypothetically determined by the subject's acuity—were associated with greater compensatory changes in the first formant. Moreover, the DIVA simulations provide an additional explanation for the observation that partial compensation was measured in Experiment 1: feedback control utilizes both auditory inputs (perturbed by the SA algorithm) and somatosensory inputs (unperturbed). That is, the presence of somatosensory feedback in the model also acts to resist changes in the feedforward commands, limiting the extent of adaptation. Thus, the DIVA model provides a plausible account of the inter-subject variation and incomplete adaptation found in this and other (Houde and Jordan, 2002; Purcell and Munhall, 2006) speech SA experiments.

Even though only tokens containing the vowel /ɛ/ received perturbed feedback, the adaptation of F1 generalized to tokens containing other vowels as well, which suggests that the subjects are not learning to modify motor commands that are specific to just the vowel that was perturbed in the SA feedback. Rather, subjects appear to have learned to modify the articulations in a way so that the adapted response can be applied globally to other vowels. Generalization is an advantageous property for speech motor planning, since speech production is normally a generative process, with each utterance being unique. Generalizing adaptation learned for one specific context to other contexts also enhances an individual's ability to rapidly modify their spoken clarity and maintain intelligibility in the face of variations in acoustic transmission conditions.[10]

At the same time, the vowels exhibited different degrees of generalization, and the vowel /i/, was the most consistently unchanged (Fig. 5). It may be necessary to hear more than a single vowel being perturbed to completely and uniformly update the vowel formant map; however, the lack of generalization for /i/ may also be explained by the possibility that, unlike other vowels, /i/ has a well-defined somatosensory target (in addition to an auditory one) that is characterized by a "saturation effect"—pressing the sides of the tongue blade against the lateral aspects of the hard palate (Perkell, 1996; Fujimura and Kakita, 1979). In DIVA, this strong somatosensory target would resist compensation to auditory perturbations since the somatosensory feedback would counteract the auditory-based compensation as soon as it tried to move the production away from the somatosensory target. Because the DIVA model learns a feedforward command for each speech sound independently, it cannot in its current form account for these generalization results, which will be used to guide future modification of the model.

While subjects' adaptive responses were expressed mainly by adjustments of the formant that was perturbed (F1), they also exhibited small changes in the second formant (F2) and even the fundamental frequency (F0), in response to the F1 perturbation. These small changes in F2 and F0 occurred in directions that were opposite of the direction of F1 changes, and would be compensatory responses if subjects perceived vowel auditory dimensions in a normalized space sensitive to relative values of formants and F0, such as the perceptual space described by Miller (1989). Simulations using the Miller space were also run (Villacorta, 2006); however, with the exception of the very small changes in F2 and F0 seen in the experimental subjects, the experimental data were better fit by versions of the model utilizing straight formant frequencies rather than the Miller space.

The quantitative simulation of human subject SA results and their relation to speaker acuity by the DIVA model adds to the list of measured speech phenomena accounted for by the model (see also Callan *et al.*, 2000, Guenther, 1994, 1995; Guenther *et al.*, 1998, 2006; Nieto-Castanon *et al.*, 2005; Perkell *et al.*, 2004a, b). The notion of tightly coupled feedforward and feedback controllers responsible for the model's ability to account for the experimental data is also a feature of other sensorimotor control CHI architectures (Wolpert and Kawato, 1998), although it is not clear if the

Villacorta *et al.*: Sensorimotor adaptation and perception of vowels
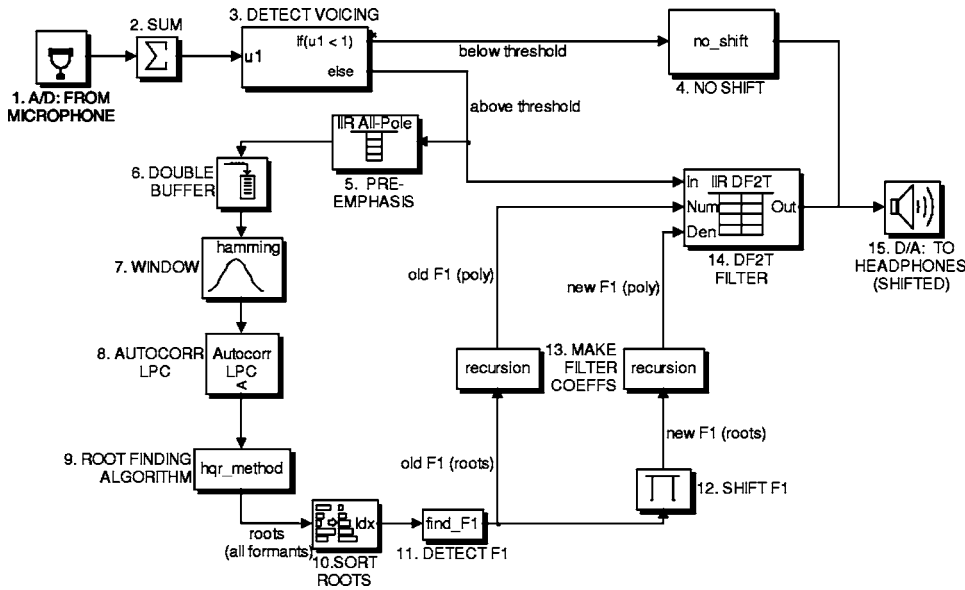
FIG. 12. Block diagram of the formant-shifting algorithm used to introduce acoustic perturbations in the SA experiment.

details of these architectures (which differ somewhat from the DIVA model) would allow quantitative fits of the current data. Specifically, we have attributed the finding that subjects exhibited a lower adaptive response in their −*feedback* vowels when compared to their +*feedback* vowels (also seen by Houde and Jordan, 2002) to the countervailing influence of unperturbed somatosensory feedback, which would be more dominant in the absence of auditory feedback. These hypotheses would be put to a more stringent test in an experiment that incorporated both auditory perturbations (Tourville *et al.*, 2004; Purcell and Munhall, 2006)—as in this study—and somatosensory perturbations (see Honda and Murano, 2003; Tourville *et al.*, 2004)—in the same experimental run.

ACKNOWLEDGMENTS

APPENDIX I. USE OF LPC COEFFICIENTS TO DETERMINE AND SHIFT F1

The following describes the signal processing that was utilized to introduce F1 shifts in nearly real time, as illustrated in Fig. 12. The DSP board receives an analog speech signal from the microphone and converts it to a digital signal, which is sent to the receiving (Rx) buffer. One of the algorithm's first functions is to calculate the sum of all values within the Rx buffer to determine its amplitude, and then determine if this value is above or below a threshold value (with signals above threshold indicating vowel production). The threshold value is set so that values below it are not sent through the formant shifting algorithm, while values above it—those within a vowel—are.

If the Rx buffer contains values from within a vowel, the signal is then pre-emphasized to improve LPC formant analysis, and coupled with the previous buffer to improve frequency resolution. A Hamming window is then applied to this double-buffered frame, followed by autocorrelation LPC analysis. The output of this operation is an 8th order polynomial (see Eq. (A1)), which can characterize up to four poles of the spectrum resulting from the analyzed speech buffer

$$A(z) = 1 + \sum_{i=1}^{8} a_i z^{-i}. \tag{A1}$$

Equation (A1) can alternatively be written with the complex roots of $A(z)$ stated explicitly, as in Eq. (A2). Written in this way, the formants of the analysis buffer are directly related to the angle θ of the complex roots. Thus, to pick out individual formants from the LPC polynomial, it is necessary to determine its complex roots, utilizing an iterative rootfinding algorithm based on the Hessenberg QR method (Press *et al.*, 2002)

$$A(z) = \prod_{i=1}^{4} (1 - c_i z^{-1})(1 - c_i^* z^{-1}),$$

where

$$c_i = r_i \cos \theta_i + j r_i \sin \theta_i,$$

$$c_i^* = r_i \cos \theta_i - j r_i \sin \theta_i. \tag{A2}$$

The roots are then sorted based on angle of the complex roots, and F1 is determined from the sorted array of roots as the lowest non-negative, nonzero root. The root related to the shifted F1 is calculated by rotating the angle of the complex root representing the original F1. A simple recursion formula is used to convert the roots of the original and shifted F1 values to polynomial coefficients. Using the new coefficients, the perturbation algorithm regenerates the speech signal with the shifted F1 value by implementing a direct-form II transposed filter (Oppenheim and Schafer, 1999) to filter data within the Rx current buffer; the original F1 is zeroed out while the perturbed F1 value is introduced simultaneously. The time-domain difference equation corresponding

to this filtering is described in Eq. (A3), with r and θ representing the magnitude and angle resulting from the complex root corresponding to the first formant

$$y[n] = x[n] - (2r \cos \theta)x[n-1] + r^2 x[n-2]$$
$$+ (2r \cos \theta')y[n-1] - r^2 y[n-2]. \quad \text{(A3)}$$

Regardless of whether the current buffer is shifted or not, the resulting speech is moved into the transfer (Tx) buffer, which is then converted back to an analog signal and sent to the output of the DSP board. The delay from the algorithm is 128 samples with sampling rate of 8000 Hz. Including a 2 ms delay resulting from anti-aliasing filtering from the analog-to-digital conversion, the overall delay in the board is 18 ms.

Aside from the threshold criterion, another criterion for shifting F1 is that the detected value had to fall within the following ranges of frequencies:

$$250 \text{ Hz} < F1 < 950 \text{ Hz} \ (male \ subjects)$$
$$\text{(A4)}$$
$$400 \text{ Hz} < F1 < 950 \text{ Hz} \ (female \ subjects).$$

F1 values below the lower limit of the window tended to be near the value of the fundamental frequency, while F1 values above the window's upper limit tended to be very close to the value of the second formant. That is, a formant value detected outside the window is likely to not be the actual F1, indicating that it should be excluded. Valid F1 values could occur outside of this range, which was a basis for rejecting the data from a preliminary subject (Villacorta, 2006).

As mentioned above, the experimental setup allowed for simultaneous recording of the input to the DSP board (no perturbation) and output of the board (with perturbation) for off-line analysis. As a validation of the effectiveness of the perturbation procedure, unperturbed and perturbed vowel formants from the same recording compared for preliminary subjects who did not hear feedback. These results showed that most of the shifted values were within 2.5% of the expected shift regardless of direction.

APPENDIX II. IMPLEMENTATION OF SUBJECT AUDITORY ACUITY IN THE DIVA MODEL

Previous versions of the DIVA model implemented a conceptualization of auditory goal regions in which auditory feedback error resulted only if the actual feedback fell outside the auditory goal region, while no feedback error resulted if the actual feedback fell within the auditory goal region (Guenther et al., 2006; Guenther, 1995; Guenther et al., 1998). Here, the discontinuity in the feedback error signal was removed by representing the actual feedback and the goal region as Gaussian distributions. The magnitude of the feedback error signal was proportional to the rectified difference between these distributions (ensuring that the feedback error signal is smaller when the actual feedback is closer to the goal), while the direction of the feedback error signal was determined by the position of the actual feedback relative to the center of the goal region (ensuring that the feedback signal compensates for the perturbation).

Speech motor planning systems of subjects of differing auditory acuity were simulated by changing the variance of the Gaussian distributions, with subjects having greater acuity possessing smaller variance values (and thus narrower goal regions). To do this, the regression line between the discrimination indices and the adaptive response indices shown in Fig. 8 was used to determine the subject's auditory region boundary size from the subject's measured ARI score. While only 13 subjects had measured discrimination index scores, all 20 subjects were simulated in the DIVA model utilizing their ARI scores and the linear relation. Had the simulation results been limited to the 13 subjects whose acuity was measured, the results would have been essentially the same, although the estimate of the true distribution of the model's productions would not have been as good.

[1]"DIVA" is an acronym for *Directions into Velocities of Articulators*; the model is so named because of its reliance on mappings that transform sensory errors into corrective motor commands.

[2]The use of insert headphones could have enhanced bone conduction of low-frequency energy, including the frequency region of F1 of the unperturbed speech signal (see Porschmann, 2000), making it possible that the subject heard a mixture of perturbed and unperturbed signals. As mentioned previously, the likelihood of the occurrence of such a confound was minimized by determining informally that the ratio of the level presented by the headphones to that of the subject's sound output, approximately 18 dB, was sufficient to mask produced vowel quality.

[3]Presentation and discussion of results from individual subjects are beyond the scope of the current report; for such details, see Villacorta (2006).

[4]Figure 3 shows a gradual increase in F1 values during the base line phase in both shift-up and shift-down subjects; this increase was especially notable in the first five epochs of the base line phase. To exclude low F1 values observed in the early part of the base line phase, the normalization shown in Eq. (1) used epochs 6–15 (an adjusted base line phase). This gradual F1 increase is discussed further in Villacorta (2006).

[5]The vowel /o/ in "pote" was also on the word list; as a diphthong, this vowel had large variations in formant values vs. time and was not included in the analysis.

[6]Subjects in Jones and Munhall (2000) produced an upward F0 shift in response to F0 shifts in their acoustic feedback, regardless of whether they were exposed to shift-up, shift-down or control protocols. Subjects exposed to the shift-down protocol increased F0 to a greater degree than the controls, while those exposed to the shift-up protocol increased in F0 to a lesser degree than the controls. When the overall increase in F0 was factored out, the subjects produced a shift in F0 that was opposite to the F0 perturbation in their auditory feedback.

[7]Initially, auditory acuity, in the form of *jnds*, were determined for each subject at three milestones: lower (at 0.85 pert), center (at 1.0 pert), and upper (at 1.15 pert). Only the results of the center milestone *jnd* determination are discussed here, as no significant findings resulted from cross-subject correlations involving the *jnd* from the two other milestones. A goodness-rating task, in which subjects were instructed to rate 41 tokens ranging from 0.7 to 1.3 pert for how well they sounded like the vowel /ɛ/, was also run as part of the perceptual experiment. Results from the goodness-rating task and the noncenter milestone *jnd* measurements are discussed in depth in Villacorta (2006).

[8]The stimulus pairs initially were separated from each other by 0.30 pert. The first four changes in separation were 0.04 pert, then subsequently by changes in separation of 0.02 pert. Once the tokens were within 0.10 pert from each other, the separation was only changed by 0.01 pert. After eight reversals (changes in direction of the staircase), the protocol terminated, and the *jnd*$_{\text{est}}$ was calculated as the median value of the last four reversals on the staircase. Two of the subjects had *jnd*$_{\text{est}}$ that were higher than the initial value set at the beginning of the staircase protocol; that is, the staircase "climbed" rather than "descended." We assumed that this was due to an initial misunderstanding of the protocol instructions, and re-ran the entire acuity experiment, including regeneration of the stimulus vowel continuum.

[9]Two subjects had adaptive response indices that were negative, indicating that they changed their productions in the same direction as the perturbation

rather than in the opposite direction. This result has been found in other auditory perturbation experiments (e.g., Burnett *et al.*, 1998). A speaker with poor acuity, who does not detect a change at all, is equally likely to have a small positive or small negative ARI. Those with good acuity have a large compensation component so their ARI is never negative. For this reason the negative adaptive responses were included in the correlation analyses rather than being removed as outliers.

[10]There is another possible explanation for generalization: vowels are learned and perhaps controlled as part of a paradigm or system. If the feature values of one vowel are modified, the system is changed in a way that could lead to changes in other vowels (Harlan Lane, personal communication).

Abbs, J. H., and Gracco, V. L. (**1984**). "Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech," J. Neurophysiol. **51**, 705–723.

Bedford, F. L. (**1989**). "Constraints on learning new mappings between perceptual dimensions," J. Exp. Psychol. Hum. Percept. Perform. **15**, 232–248.

Bhushan, N., and Shadmehr, R. (**1999**). "Computational nature of human adaptive control during learning of reaching movements in force fields," Biol. Cybern. **81**, 39–60.

Blakemore, S. J., Goodbody, S. J., and Wolpert, D. M. (**1998**). "Predicting the consequences of our own actions: The role of sensorimotor context estimation," J. Neurosci. **18**, 7511–7518.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice F0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**, 3153–3161.

Callan, D. E., Kent, R. D., Guenther, F. H., and Vorperian, H. K. (**2000**). "An auditory−feedback based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system," J. Speech Lang. Hear. Res. **43**, 721–736.

Fujimura, O., and Kakita, Y. (**1979**). "Remarks on quantitative description of lingual articulation," in *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Öhman (Academic, San Diego), pp. 17–24.

Guenther, F. H. (**1995**). "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production," Psychol. Rev. **102**, 594–621.

Guenther, F. H. (**1994**). "A neural network model of speech acquisition and motor equivalent speech production," Biol. Cybern. **72**, 43–53

Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M., and Perkell, J. S. (**1999b**). "Articulatory tradeoffs reduce acoustic variability during American English /r/ production," J. Acoust. Soc. Am. **105**, 2854–2865.

Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (**2006**). "Neural modeling and imaging of the cortical interactions underlying syllable production," Brain Lang **96**, 280–301.

Guenther, F. H., Hampson, M., and Johnson, D. (**1998**). "A theoretical investigation of reference frames for the planning of speech movements," Psychol. Rev. **105**, 611–633.

Guenther, F. H., Husain, F. T., Cohen, M. A., and Shinn-Cunningham, B. G. (**1999a**). "Effects of categorization and discrimination training on auditory perceptual space," J. Acoust. Soc. Am. **106**, 2900–2912.

Guenther, F. H., Nieto-Castanon, A., Ghosh, S. S., and Tourville, J. A. (**2004**). "Representation of sound categories in auditory cortical maps," J. Speech Lang. Hear. Res. **47**, 46–57.

Honda, M., and Murano, E. (**2003**). "Effects of tactile and auditory feedback on compensatory articulatory response to an unexpected palatal perturbation," Proceedings of the Sixth Speech Production Seminar, Sydney, Australia, Dec. 7–10, 2003.

Houde, J. F., and Jordan, M. I. (**1998**). "Sensorimotor adaptation in speech production," Science **279**, 1213–l216.

Houde, J. F., and Jordan, M. I. (**2002**). "Sensorimotor adaptation of speech I: Compensation and adaptation," J. Speech Lang. Hear. Res. **45**, 295–310.

Jones, J. A., and Munhall, K. G. (**2000**). "Perceptual calibration of F0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**, 1246–1251.

Kawahara, H. (**1993**). "Transformed auditory feedback: Effects of fundamental frequency perturbation," J. Acoust. Soc. Am. **94**, 1883–1884.

Kawato, M., and Gomi, H. (**1992**). "The cerebellum and VOR/OKR learning models," Trends Neurosci. **15**, 445–453.

Lane, H. L., and Tranel, B. W. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Lang. Hear. Res. **14**, 677–709.

Lindblom, B. E. F., Lubker, J. F., and Gay, T. (**1979**). "Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation," J. Phonetics **7**, 147–161.

Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A User's Guide*, 2nd ed. (Lawrence Erlbaum, Mahwah, NJ).

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. 131–149.

Markel, J. D., and Gray, A. H. (**1976**). *Linear Prediction of Speech* (Springer-Verlag, New York).

Max, L., Wallace, M. E., and Vincent, I. (**2003**). "Sensorimotor adaptation to auditory perturbations during speech: Acoustic and kinematic experiments," *Proceedings of the 15th Internatiional Congress of Phonetic Sciences*, Barcelona, pp. 1053–1056.

Miller, J. D. (**1989**). "Auditory-perceptual interpretation of the vowel," J. Acoust. Soc. Am. **85**, pp. 2114–2134.

Newman, R. S. (**2003**). "Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report," J. Acoust. Soc. Am. **113**, 2850–2860.

Nieto-Castanon, A., Guenther, F. H., Perkell, J., and Curtin, H. D. (**2005**). "A modeling investigation of articulatory variability and acoustic stability during American English /r/ production," J. Acoust. Soc. Am. **117**, 3196–3212.

Oppenheim, A. V., and Schafer, R. W. (**1999**). *Discrete-Time Signal Processing*, 2nd ed. (Prentice–Hall, Upper Saddle River, NJ).

Perkell, J. S. (**1996**). "Properties of the tongue help to define vowel categories: Hypotheses based on physiologically oriented modeling," J. Phonetics **24**, 3–22.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E. S., and Tiede, M. (**2004a**). "The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts," J. Acoust. Soc. Am. **116**, 2338–2344.

Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., and Marrone, N. (**2004b**). "The distinctness of speakers' /s/−/∫/ contrast is related to their auditory discrimination and use of an articulatory saturation effect," J. Speech Lang. Hear. Res. **47**, 1259–1269.

Porschmann, C. (**2000**). "Influences of bone conduction and air conduction on the sound of one's own voice," Acta Acust. (Beijing) **86**, 1038–1045.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (**2002**). *Numerical Recipes in C*, 2nd ed. (Cambridge University Press, Cambridge).

Purcell, D. W., and Munhall, K. G. (**2006**). "Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation," J. Acoust. Soc. Am. **120**, 966–977.

Savariaux, C., Perrier, P., and Orliaguet, J. P. (**1995**). "Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production," J. Acoust. Soc. Am. **98**, 2428–2842.

Tourville, J. A., Guenther, F. H., Ghosh, S. S., and Bohland, J. W. (**2004**). "Effects of jaw perturbation on cortical activity during speech production," J. Acoust. Soc. Am. **116**, 2631(A).

Vallabha, G. K., and Tuller, B. (**2002**). "Systematic errors in the formant analysis of steady-state vowels," Speech Commun. **38**, 141–160.

Villacorta, V. M. (**2006**). "Sensorimotor adaptation to perturbations of vowel acoustics and its relation to perception," Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Welch, R. B. (**1978**). *Perceptual Modification: Adapting to Altered Sensory Environments* (Academic, New York).

Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (**1995**). "Are arm trajectories planned in kinematic or dynamic coordinates? An adaptation study," Exp. Brain Res. **103**, 460–470.

Wolpert, D. M., and Kawato, M. (**1998**). "Multiple paired forward and inverse models for motor control," Neural Networks **1217**, 1–13.

Xu, Y., Larson, C. R., Bauer, J. J., and Hain, T. C. (**2004**). "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," J. Acoust. Soc. Am. **116**, 1168–1178.

Yates, A. J. (**1963**). "Delayed auditory feedback," Psychol. Bull. **60**, 213–232.

A role for the second subglottal resonance in lexical access[a]

Steven M. Lulich[b]
Speech Communication Group, MIT, and Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts 02139

Asaf Bachrach
Department of Linguistics and Philosophy, MIT, Cambridge, Massachusetts 02139

Nicolas Malyska
MIT Lincoln Laboratory, Lexington, Massachusetts 02420 and Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts 02139

Acoustic coupling between the vocal tract and the lower (subglottal) airway results in the introduction of pole-zero pairs corresponding to resonances of the uncoupled lower airway. If the second formant (F2) passes through the second subglottal resonance a discontinuity in amplitude occurs. This work explores the hypothesis that this F2 discontinuity affects how listeners perceive the distinctive feature [back] in transitions from a front vowel (high F2) to a labial stop (low F2). Two versions of the utterances "apter" and "up there" were synthesized with an F2 discontinuity at different locations in the initial VC transition. Subjects heard portions of the utterances with and without the discontinuity, and were asked to identify whether the utterances were real words or not. Results show that the frequency of the F2 discontinuity in an utterance influences the perception of backness in the vowel. Discontinuities of this sort are proposed to play a role in shaping vowel inventories in the world's languages [K. N. Stevens, J. Phonetics **17**, 3–46 (1989)]. The results support a model of lexical access in which articulatory-acoustic discontinuities subserve phonological feature identification. © 2007 Acoustical Society of America.
[DOI: 10.1121/1.2772227]

I. INTRODUCTION

The quantal theory of the relation between phonological features and their acoustic and articulatory correlates suggests that nonlinearities in articulatory-acoustic space define the boundaries between plus and minus values of some features (Stevens, 1972, 1989, 1998). One series of discontinuities results from the acoustic coupling between the lower (subglottal) airway and the upper (vocal tract) airway. Due to the location of the voice source (at the glottis) between these two airways, the resonances of the lower airway introduce pole-zero pairs into speech spectra. The zeros are located at the frequencies of the lower airway resonances and the poles are somewhat offset toward higher frequencies, depending on the coupling of the lower and upper airways. Since the subglottal system is relatively unaltered during speech, each subglottal pole-zero pair forms a quasistationary benchmark in the acoustic signal that contributes little spectral energy at all frequencies except near its center frequency (i.e., the frequency between the pole and the zero at which the total energy is 0 dB; cf. Fig. 1.) Near the center frequency, the amplitude of the spectrum is significantly affected by the

pole-zero pair. Thus, over the narrow frequency band defined by the pole-zero pair, spectral amplitude is roughly discontinuous.

The second resonance of the lower airway has been found from direct (e.g., input impedance) measurements to be near 1400 Hz for males (Cranen and Boves, 1987; Ishizaka *et al.*, 1976), and measurements of the corresponding pole-zero pairs in speech signals have been consistent with this (Fant *et al.*, 1972; Hanson, 1996; Hanson and Stevens, 1995; Klatt and Klatt 1990). The second subglottal resonance lies near the frequency which divides front vowels from back vowels: front vowels have a high F2 ($>$1400 Hz) whereas back vowels have a low F2 ($<$1400 Hz). This observation was first made by Stevens (1998) and followed up in an experiment by Chi and Sonderegger (2004) (cf. Chi and Sonderegger, 2007), in which they recorded the speech signal simultaneously with the signal from an accelerometer placed above the sternal notch, the formants of which are the subglottal resonances (cf. Cheyne, 2002; Stevens *et al.*, 1975). They found that individual speakers of American English produce front and back vowels with F2 values consistently higher or lower (respectively) than the second subglottal resonances of the same speakers.

In natural speech, second formant transitions across the second subglottal resonance occur frequently and often show a discontinuous-amplitude effect (cf. Chi and Sonderegger, 2007). In addition to the amplitude discontinuity, there is a
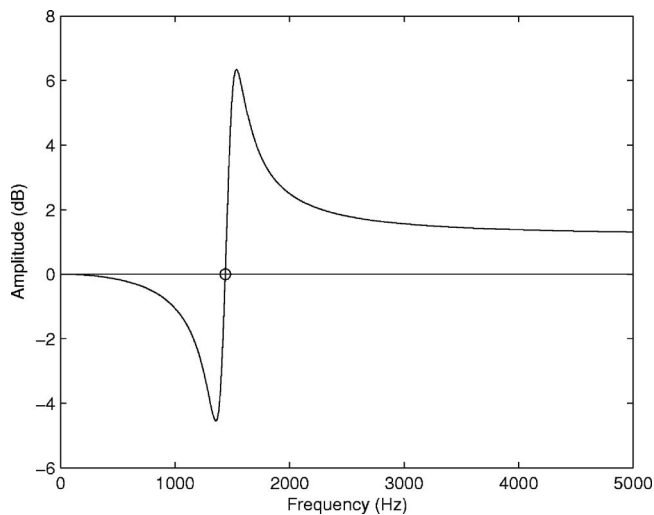
FIG. 1. A pole-zero pair. The pole has frequency 1500 Hz and the zero has frequency 1400 Hz. Both the pole and zero have a bandwidth of 150 Hz. The circle indicates the center frequency of the pole-zero pair.

FIG. 2. A schematic diagram of the gating paradigm used in the experiment.

discontinuity in F2 frequency. This is due to the fact that F2, since it is itself a pole, cannot in fact cross the subglottal pole (see Stevens, 1998, pp. 299–303).

We hypothesized that the discontinuous-amplitude effect of the second subglottal pole-zero pair plays a role in the perception of vowels in a vowel-consonant transition. In particular, we tested whether manipulations of the zero frequency affected subjects' perception of vowel quality in a vowel-initial utterance. In order to determine whether such an effect might occur in normal speech perception, the experiment was constructed in such a way that the subjects were performing a linguistic task. We predict that the discontinuous-frequency (as opposed to amplitude) effect will also influence vowel perception, but we leave that question for future research.

The distinction between front and back vowels has received several acoustic treatments (Chistovich, 1985; Ito *et al.*, 2001; Stevens, 1989, 1998; Syrdal and Gopal, 1986), most of which rely in some way on the interaction of the second formant with other formants. Stevens (1998) suggests that it is the interaction of the second formant with the second subglottal resonance that defines the front/back vowel categories. We tested this hypothesis by manipulating the frequency of a zero, but such a manipulation also affects the frequencies and amplitudes of the other formants. In order to determine whether these affects could lead to results consistent with any of the other hypotheses, we performed a post-hoc analysis of the stimuli. It was found that these other hypotheses were not consistent with the results (see Sec. IV).

II. METHODS

A. The stimuli: An overview

A vowel [æ] was copy-synthesized and simplified slightly in order to manipulate a single parameter (the frequency of the second subglottal resonance) systematically. Two versions of the synthetic vowel were made, in which the second subglottal resonance was either 1300 or 1500 Hz. Two naturally produced utterances ("up there" and "apter")
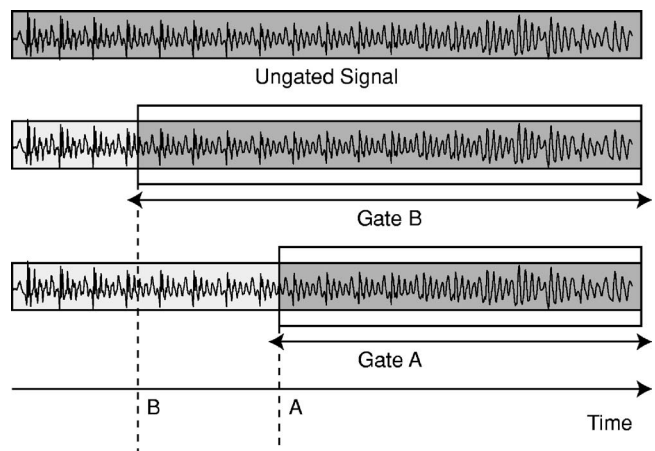
were recorded by an adult male native speaker of standard American English with a second subglottal resonance at 1384 Hz.[1] The initial vowel in each was deleted and the synthetic vowel [æ] inserted in its place. There were thus four basic utterances ("[æ]pter" with a 1500 Hz zero, "[æ]pter" with a 1300 Hz zero, "[æ]p there" with a 1500 Hz zero, and "[æ]p there" with a 1300 Hz zero). The utterances were then systematically truncated (gated; cf. Grosjean, 1996; the concept is illustrated schematically in Fig. 2) eight times within the synthetic vowel to form 32 stimuli, such that each stimulus contained some portion of the synthetic vowel followed by the naturally produced speech.

Natural utterances (by the same speaker) of the words "upter" and "up there" were used to create fillers. Each word was gated four times, yielding eight fillers. Both filler words began with the back vowel [ʌ] so that some words with long gates would be clearly perceived as back vowels, preventing subjects from relying on gate durations when making judgments. Pilot experiments indicated that both the original utterances and those in which the synthetic vowel was inserted were intelligible.

B. The synthetic vowel

We copy-synthesized the initial vowel in "apter" while making some reasonable simplifications. In the natural vowel, F2 began near 1800 Hz and fell to near 1200 Hz before the labial closure. In the synthetic vowel, F2 fell linearly from 1800 to 1200 Hz in 145 ms. In both the natural and synthetic vowels, F1 rose slightly over 110 ms and then fell toward the labial closure. F3 fell slightly over the duration of the natural vowel, and in the synthetic vowel it fell linearly from 2420 to 2400 Hz in 145 ms. Spectrograms of the copied (natural) and synthetic vowels are given in Figs. 3 and 4, respectively.

In the natural vowel, the effects of the second subglottal resonance on F2 were visible in the spectrogram, with a decrease in F2 amplitude near 1400 Hz. The visual effect is more striking in the zoomed wave form, which was bandpass-filtered between 1000 and 2000 Hz. The arrows in the spectrogram and wave form are time-aligned, and point to a pulse that is weaker than its neighbors.

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Lulich *et al.*: Subglottal resonances in speech perception   2321
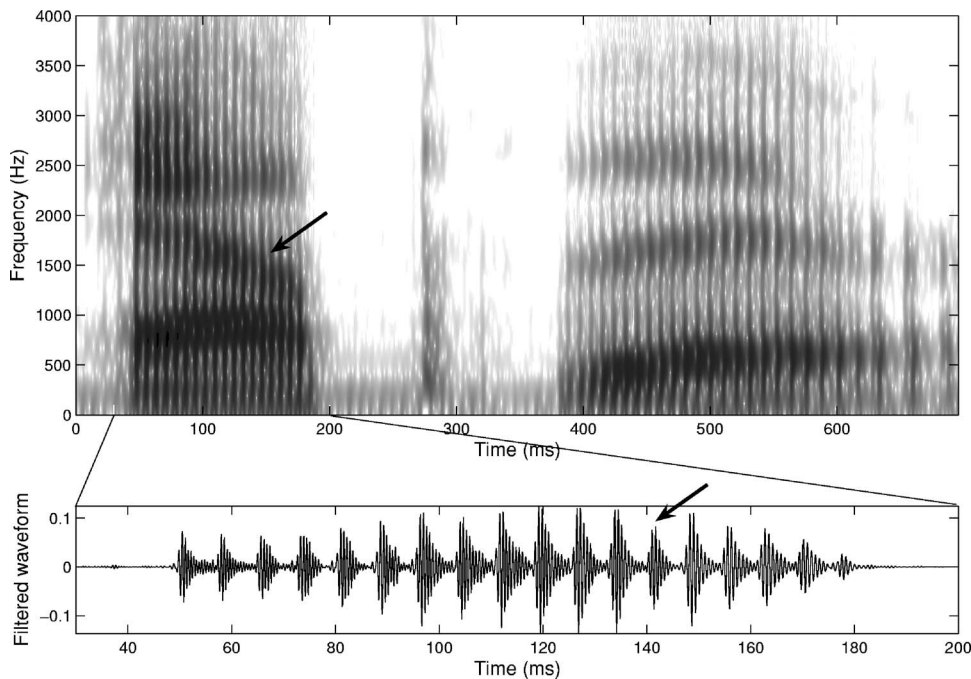
FIG. 3. Spectrogram and zoomed bandpass-filtered wave form of the naturally produced vowel [æ] in the nonword "ap there." The arrows indicate a slight decrease in F2 amplitude as it crosses the second subglottal zero. This decrease is more clearly seen in the wave form than in the spectrogram.

In the synthetic vowel, we simplified the pole-zero pair and replaced it with an effective zero. The elimination of the subglottal pole allowed us to control the stimuli and interpret the results with greater precision. When the effective zero is synthesized at 1500 Hz, the amplitude of F2 near this frequency is diminished (indicated by the large arrow in Fig. 4).

Since the synthesizer we used required pole-zero pairs (Klatt and Klatt, 1990), our effective zero was synthesized by a zero and a pole at the same frequency. The pole had a bandwidth of 500 Hz, and the zero had a bandwidth of 100 Hz. Our synthetic F2 had a bandwidth of 80 Hz, and F1 had a bandwidth of 40 Hz. Figure 5 shows the spectrum of the effective zero in relation to its component pole-zero parts, and in relation to F2 at the same frequency.

We synthesized the zero at two different frequencies, creating the two conditions for our experiment. In one condition the zero was synthesized at 1300 Hz, and in the other condition it was synthesized at 1500 Hz. In an additional control condition, there was no zero in the synthetic vowel. Because F2 fell linearly from 1800 to 1200 Hz, it crossed the zero relatively late in the vowel in the 1300 Hz condition, and relatively early in the 1500 Hz condition. We hypothesized that the stimuli with a 1300 Hz zero would be perceived as containing an initial front vowel more often than the corresponding stimuli with a 1500 Hz zero, since F2 is higher than the zero for a longer portion of the vowel transition in the 1300 Hz condition than in the 1500 Hz condition.

C. Gating

The stimuli were gated eight times in the synthetic vowel. Consecutive gates were separated by a single pitch period, and located at the wave form zero-crossing just before the initial rise at the beginning of the pitch period. The first gate was located at the beginning of the 7th pitch period from the end of the vowel (indicated by the bottom right arrow in Fig. 4); the eighth gate was located at the beginning of the 14th pitch period from the end of the vowel (indicated by the bottom left arrow in Fig. 4). There were 20 pitch
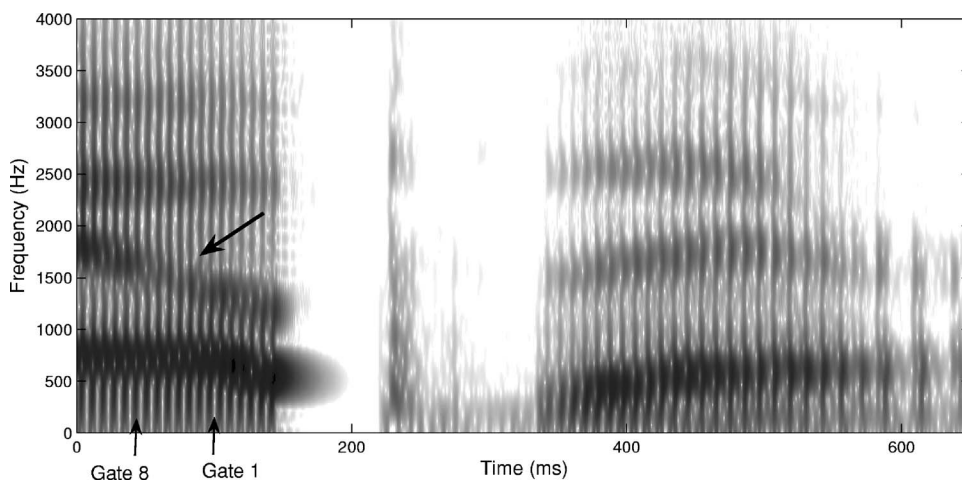
FIG. 4. The structure of the stimuli (for the nonword "ap there"). The initial vowel was synthetic, but the rest of the stimulus was naturally produced speech. The large arrow indicates the location of the visible effect of the zero on F2 (1500 Hz in this case). The left bottom arrow indicates the longest (8th) gate, and the right arrow indicates the shortest (1st) gate. Each stimulus contained the portion of the utterance beginning at one of the gates and ending at the end of the utterance. Compare gates 1 and 8 with gates A and B in Fig. 2.
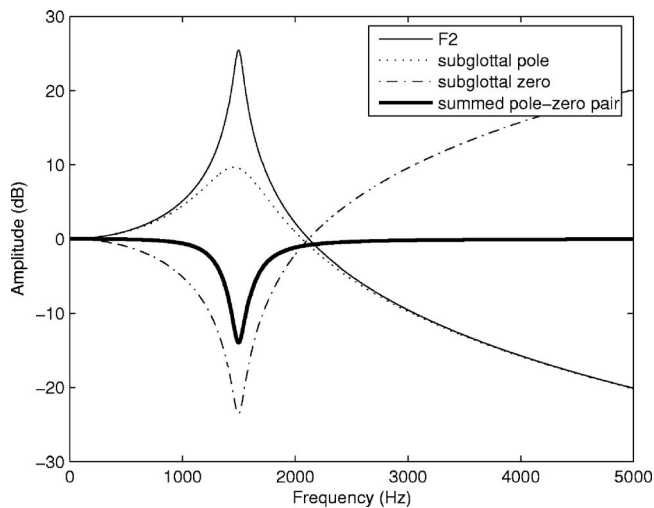
FIG. 5. The effective zero was synthesized by placing a pole and a zero at the same frequency but with different bandwidths. Bandwidths were: 80 Hz for F2; 100 Hz for the zero; 500 Hz for the pole.

FIG. 6. Identification curves for the feature [back] with standard deviation for the 1300 and 1500 Hz conditions, averaged across subjects.

periods in the whole vowel. The range of pitch periods that were gated were chosen so that they encompassed the region of the transition from perception of a back vowel to perception of a front vowel as more and more of the vowel was presented.

The filler words were gated four times in the same way as the stimuli, except there were two pitch periods between each consecutive gate. The first gate was located at the beginning of the 7th pitch period from the end of the vowel, and the fourth gate was located at the beginning of the 13th pitch period from the end of the vowel.

The fundamental frequency during the synthetic vowel was not constant, and hence individual pitch periods did not have the same duration. In the range of pitch periods that were gated, periods averaged 8 ms; and varied by less than 1 ms.

The longest gates (those with the earliest onsets) contain parts of the vowel in which the second formant crosses the effective zero; the shortest gates (those with the latest onsets) contain only those parts of the vowel in which the second formant does not cross the effective zero. Stimuli with the longest gates were expected to be perceived as containing an initial front vowel, and stimuli with the shortest gates were expected to be perceived as containing an initial back vowel.

D. Procedure

The experiment was repeated twice. For the first repetition, the 1300 and 1500 Hz conditions were tested. The 1500 Hz and control (no zero) conditions were tested in the second repetition.

In an initial training phase of the experiment, the 32 stimuli and 8 fillers were presented to the subjects in random order with no feedback. Following a 30 s pause, the 32 stimuli and 8 fillers were presented 15 times each, in random order. Thus, there were 40 training stimuli and 560 experimental stimuli, yielding a total of 600 stimuli. After half of the 560 experimental stimuli were presented, there was another 30 s pause.

In each trial, the stimulus was presented twice with a 750 ms interval between the onset of each presentation. Subjects performed a two alternative forced choice lexical identification task in each trial. There were two sets of alternative choices, appropriate to the stimulus, as follows: one set was "upter" versus "not 'upter'," and the other set was "up there" versus "not 'up there.'," Responses were not timed, and subjects were informed at the beginning that they could pace themselves. Most subjects completed the experiment in about 30 min.

Subjects listened to the stimuli in a sound treated booth, wearing a pair of Cyber Acoustics HE-200 headphones. The volume was fixed by the experimenter at a moderately low level.

E. Subjects

Eleven subjects participated in the first experiment. The data from one subject were not included in the analysis because that subject failed to perceive any of the "apter" stimuli as beginning with a back vowel. Eight subjects participated in the second experiment. The data from one subject were not included in the analysis because that subject did not follow the directions. Two of the subjects participated in both experiments. All 15 subjects were native speakers of American English, and graduate students at MIT. They were naive to the purpose of the experiment. There were five males and five females for the first experiment, and two males and five females for the second experiment.

III. RESULTS

A. Experiment 1: 1300 Hz vs 1500 Hz

Figure 6 shows the results of the first experiment, averaged across subjects and across the two sets of stimuli ("apter" and "up there"). Plotted along the abscissa are the different gates, with longer gates toward the left and shorter gates toward the right. The percentage of stimuli ($n=30$) reported to contain a back vowel are plotted along the ordi-

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Lulich *et al.*: Subglottal resonances in speech perception   2323
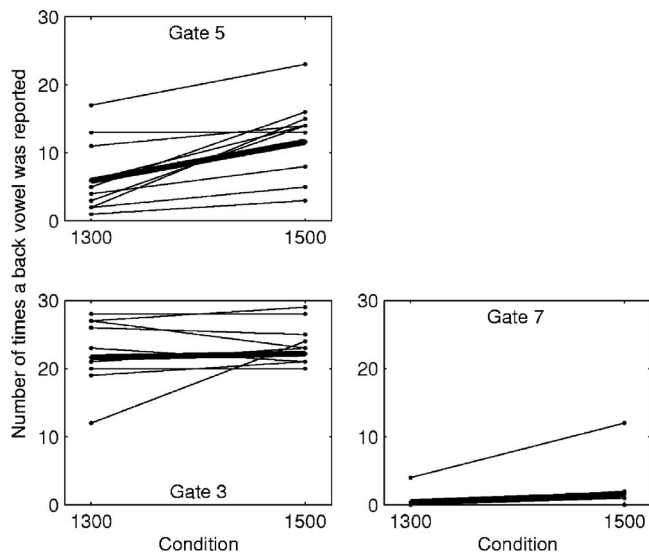
FIG. 7. Individual subjects' responses to three pairs of stimuli (thick lines are averages). In gates 3 and 7 subject responses in the 1300 and 1500 Hz conditions were similar; in gate 5 the responses differed, with more back vowels being reported in the 1500 Hz condition.

nate. Thus, for the longest gates, few stimuli were reported to contain a back vowel, whereas most stimuli were reported to contain a back vowel for the shortest gates.

For the intermediate gates (gates 4–6), there is a significant difference between the 1300 and 1500 Hz conditions (paired t-test, $p < 0.01$ for gate 5, $p < 0.03$ for gate 4, $p < 0.1$ for gate 6, $p > 0.1$ for all others). In the 1500 Hz condition more stimuli were reported to contain a back vowel than in the 1300 Hz condition. Standard deviation bars are shown in order to illustrate the variation across subjects. All subjects showed similar response patterns (cf. Fig. 7, which shows the gate by gate breakdown of subjects' responses for gates 3, 5, and 7). For intermediate gates, each subject reported more words to contain a back vowel in the 1500 Hz condition than in the 1300 Hz condition. A point-biserial correlation coefficient of $r = 0.5$ was obtained for the 5th gate, in which the difference between the two conditions was greatest. For the 3rd and 7th gates, the point-biserial correlation coefficient was less than 0.25.

B. Experiment 2: No zero vs 1500 Hz

Figure 8 shows the results of the second experiment, averaged across subjects and across the two sets of stimuli ("apter" and "up there"). The general pattern is identical to that in the first experiment.

For most gates (gates 3–8), there is a significant difference between the control (no zero) and 1500 Hz conditions (paired t-test, $p < 0.01$ for gates 4–6 $p < 0.05$ for gates 3 and 7–8, $p > 0.1$ for gates 1 and 2). In the 1500 Hz condition more stimuli were reported to contain a back vowel than in the control condition. Standard deviation bars are shown in order to illustrate the variation across subjects. All subjects showed similar response patterns.

Across the two experiments, there was a significant difference found between the 1300 Hz and control conditions
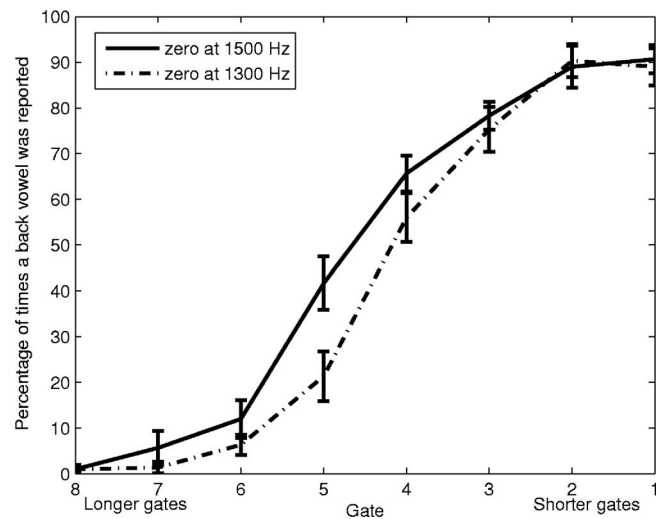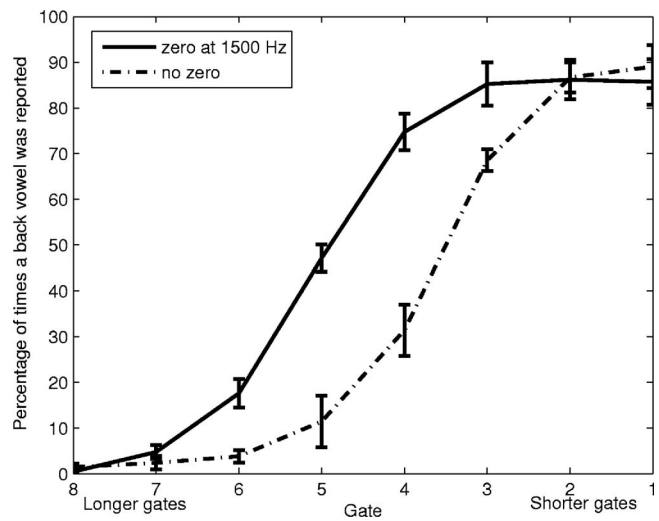

FIG. 8. Identification curves for the feature [back] with standard deviation for the control (no zero) and 1500 Hz conditions, averaged across subjects.

for gate 4 (two-sample t-test, $p < 0.02$), and no significant difference was found between the 1500 Hz conditions in the two experiments ($p > 0.1$).

IV. DISCUSSION

The results indicate that the perception of vowel backness in the context of a spoken utterance can be altered as a result of manipulating either the presence or the frequency of an effective acoustic zero in the vowel spectrum. Thus far we have followed Stevens (1998) and Chi and Sonderegger (2004) in suggesting that the boundary between front and back vowels is defined by the location of the second subglottal resonance. The results are consistent with this view. However, the presence of the effective acoustic zero in the vowel causes several changes to the overall spectrum of the vowel, including the apparent frequency and amplitude of the second formant, and the ratio of these parameters to those of the third formant. Thus, it is possible that the cause of the altered percept of vowel backness may be due to one of these other, secondary factors, rather than to the frequency of the acoustic zero per se. In this section, we will briefly review three alternative explanations for our results. Ultimately, none of these alternative explanations appears to be plausible.

A. Alternative 1: F3-F2 bark difference

Several studies have found that front and back vowels may be distinguished from each other by referring to the frequency difference between the second and third formants (Chistovich, 1985; Syrdal and Gopal, 1986). Specifically, if F2 and F3 are separated by less than roughly 3.5 bark, the vowel may be classified as a front vowel; if F2 and F3 are separated by more than 3.5 bark, the vowel may be classified as a back vowel. Since our stimuli had a descending F2 during the vowel, the difference between F2 and F3 began less than 3.5 bark and ended greater than 3.5 bark. It is interesting to note that the region in which the results of the 1300 and 1500 Hz conditions differed was when F2 and F3 were roughly 3.5 bark separated. This was also the region
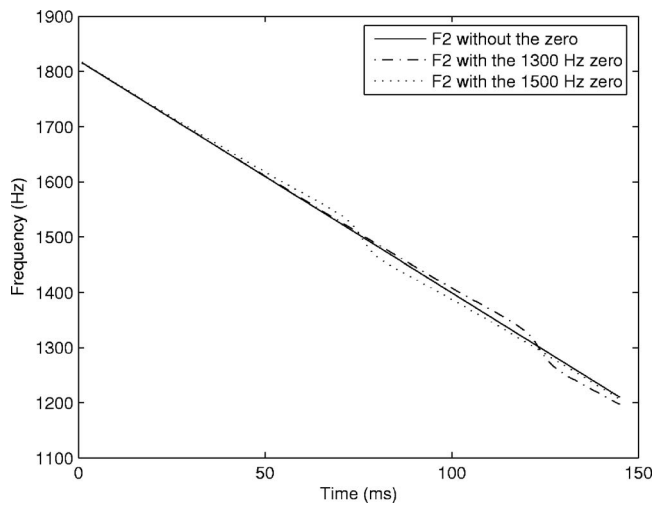
FIG. 9. Input F2 and apparent F2 in the 1300 and 1500 Hz conditions. The thin solid line is the input F2, descending linearly. The dash-dot line is the apparent F2 in the 1300 Hz condition. The dotted line is the apparent F2 in the 1500 Hz condition.

FIG. 10. Amplitude of apparent F3 and F2 in the 1300 and 1500 Hz conditions, and in the control (no zero) condition.

forming the boundary between front and back vowel percepts more generally. Thus, it may be the case that the 3.5 bark separation between F2 and F3 forms at least a crude boundary between front and back vowels, which may be further defined with respect to the second subglottal resonance. Alternatively, it is also possible that the 3.5 bark separation between F2 and F3 functions as a boundary between front and back vowels precisely because 3.5 bark less than F3 is roughly the location of the second subglottal resonance in natural speech.

Regardless of the fact that the transition region in the identification curves of Figs. 6 and 8 coincides with a roughly 3.5 bark separation between F2 and F3, it is important to note that the percept of vowel backness was manipulated within this region. It is possible that this effect could be accounted for in terms of the separation between F2 and F3, since the actual frequency of the output peak amplitude near F2 is not identical to the input F2 value (as shown in Fig. 9). However, the differences between the 1300 and 1500 Hz conditions are very small (on the order of 1% or smaller), whereas the 3.5 bark boundary hypothesis has never been demonstrated to this degree of precision. Therefore the 3.5 bark boundary hypothesis would not predict the results reported here, either for the 1500 Hz vs 1300 Hz conditions, or for the 1500 Hz vs control (no zero) conditions.

B. Alternative 2: Center of gravity (F2′)

Closely related to the 3.5 bark boundary hypothesis is the finding that closely spaced formants may be analyzed as a single excitation distribution by the auditory system, with a frequency-domain center of gravity dependent not only on the frequencies of the formants but also on their relative amplitudes (Chistovich, 1985). Center of gravity effects occur only when the formants in question are within about 3.5 bark of each other, close enough together that they may be integrated into a single distribution. Thus, if the lower formant has a relatively high amplitude, the center of gravity will be biased toward low frequencies; if the lower formant
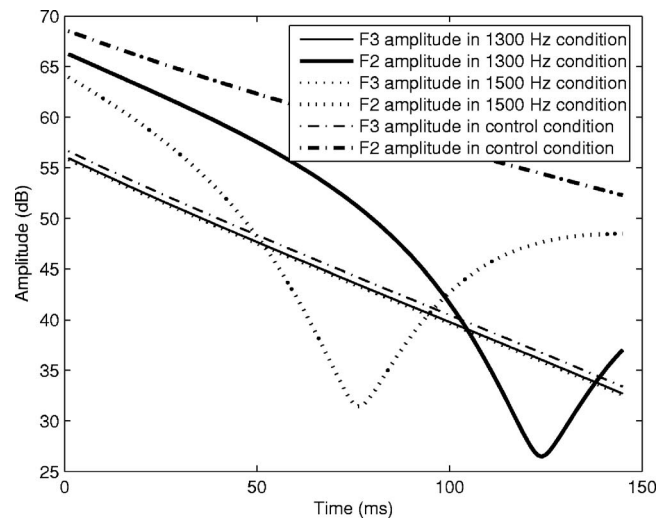
has a relatively low amplitude, the center of gravity will be biased toward high frequencies. If F2 and F3 are the two formants being integrated into a single center of gravity measure (F2′), vowels with a high F2 amplitude will be more likely to be perceived as back vowels, since the center of gravity is lower; vowels with a low F2 amplitude will be more likely to be perceived as front vowels, since the center of gravity is higher.

The effective acoustic zero introduced in the synthesis of our vowel stimuli affected the amplitudes of F2 and F3 (as shown in Fig. 10). The differences in F2 amplitude between the 1300 and 1500 Hz conditions approach 20 dB, whereas the difference in F3 amplitude is negligible. In the early part of the vowel ($<$100 ms), the F2 amplitudes in the two conditions would actually predict the opposite of the results we obtained, if the center of gravity effect were the cause. Here, the F2 amplitude is smaller in the 1500 Hz condition than in the 1300 Hz condition, which should lead to a higher center of gravity (F2′) in the 1500 Hz condition, resulting in more front vowel reports. We found that there were more back vowel reports in the 1500 Hz condition than in the 1300 Hz condition. In the later part of the vowel ($>$100 ms), the formants are far enough apart that they should not be analyzed as a single distribution. The center of gravity hypothesis is therefore ruled out for the 1500 Hz vs 1300 Hz conditions. It is also ruled out for the 1500 Hz vs control (no zero) conditions by the same reasoning.

C. Alternative 3: Spectral tilt

Recent studies have explored the possibility that spectral tilt plays a role in distinguishing front from back vowels. Specifically, front vowels should have a smaller spectral tilt than back vowels, because a high second formant will boost the amplitudes of all higher formants. However, this effect has been demonstrated only in steady-state vowels (Ito et al., 2001), and was not found for time-varying vowels (Kiefte and Kluender, 2005). Since our synthetic vowel stimuli were time-varying, it is therefore unlikely that the spectral tilt hypothesis can account for the results.[2]

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Lulich *et al.*: Subglottal resonances in speech perception   2325

D. Auditory detectability of the zero

We have suggested that an amplitude discontinuity in F2 arising from the interaction of F2 with a zero is perceptually salient. Moreover, we have suggested that such an interaction involving the zero associated with the second subglottal resonance plays a role in identifying vowel quality in a vowel-consonant transition. This raises two questions. First, is there independent evidence that the auditory system is sensitive to discontinuities as small as that caused by the second subglottal resonance in natural speech? Second, could such discontinuities at frequencies other than the second subglottal resonance similarly affect vowel quality perception?

The effective zero used in the synthesis of our stimuli had a depth of roughly 14 dB and its effect on the amplitude of F2 was strongest over a time span less than 50 ms. Both the depth of the zero and the time span of its amplitude effect are larger than might be expected in real speech. However, zeros with depths as little as 2 dB have been reported to be perceptually salient (Flanagan, 1957; Malme, 1959, 1960), and the change in amplitude between a pole and a zero can be as much as 10 dB, even when their bandwidths are reasonably large (cf. Fig. 1). Furthermore, it has been shown that abrupt decreases in amplitude lasting as little as 3 ms are detectable in amplitude-modulated sinusoidal signals (Ross and Pantev, 2004). Although it is difficult to compare results from such psychoacoustic studies directly with the results presented here, we believe that our interpretation is reasonably consistent with what is known about auditory processing of sound. That is, the presence of an acoustic zero can plausibly have an effect on vowel perception.

In addition to the detection of the decrease in F2 amplitude as it passes through the zero, it is possible that the presence of the zero has an effect on perception even when F2 does not pass through it. That is, if a portion of the vowel is played to subjects such that F2 always remains below the zero, it is still possible that the frequency of the zero could have an effect on the perception of the vowel, especially when the formant is moving. Future psychoacoustic work in this area will be crucial for further defining the effect of the subglottal resonances in speech perception.

The fact that the relation of F2 both to the second subglottal resonance and to F3 provides a means of distinguishing front and back vowels suggests that the relation between F3 and the subglottal resonance might also be important. If, for instance, F3 is raised, or if the second subglottal resonance is lowered, so that the second subglottal resonance and F3 are separated by more than 3.5 bark, the effect of the second subglottal resonance on vowel quality perception might decrease. Conversely, if F3 is lowered, or if the second subglottal resonance is raised, so that they are separated by less than 3.5 bark, the perceptual effect of the second subglottal resonance might similarly decrease. Such a relationship between F3 and the second subglottal resonance could explain how perception of vowel quality in noise or degraded speech might still be possible, for even if the subglottal resonance is masked by noise F3 could be used by listeners either as a parallel strategy for vowel quality perception, or to construct a percept of the second subglottal resonance itself, analogous to the well-known phenomenon of pitch perception with a "missing fundamental."

E. Quantal theory

The quantal theory (Stevens, 1972, 1989, 1998) hypothesizes that at least some speech sound contrasts are defined by discontinuities in articulatory-acoustic space. In the case of front and back vowels, a discontinuity is supplied by the effect of the second subglottal resonance (Stevens, 1998, pp. 299–303).

Classically, quantal theory has been understood only in spectral terms, i.e., as a phenomenon in the frequency domain. A second theory developed by Stevens (2002) posits that listeners are sensitive to acoustic "landmarks," which in effect "tell" the listener where to look for cues to a certain feature value. The determination of the particular feature value is supposed to be made in light of its spectral position within quantal space. In fact, the quantal theory and the theory of landmarks are two sides of the same coin—landmarks in the spectrographic domain are to be expected as a result of temporal transitions from one side of a quantally defined spectral contrast to the other. Thus far, thinking about landmarks has been restricted to fairly wide-band discontinuities in the temporal structure of speech and are supposed to be caused by changes in manner, or articulator-free, features.

The design and results of our experiment were such that the presence of a narrow-band acoustic landmark, i.e., the discontinuity in F2 as it crossed the subglottal zero, or a more complicated auditory process related to the detection of this landmark, affected listeners' perception of a front vowel transitioning to a labial stop. In this sense we propose an extension of the theory of landmarks to include such narrow-band discontinuities caused by changes in place, or articulator-bound, features. Our results further suggest that the theory of landmark detection (and therefore also of quantal theory) has real psycholinguistic substance. It is expected that future experiments, which will include a more realistic pole-zero pair rather than an effective zero only, will show larger effects, since two sources of discontinuity will then be present.

Earlier versions of quantal theory (Stevens, 1972, 1989), suggesting that stable regions in articulatory-acoustic space were preferred for defining phonological categories, hypothesized that such stable regions could be obtained when two formants are near together (as in the case of back vowels in which F1 and F2 are close, or of front vowels in which F2 and F3 are close). The same should therefore be true when the second formant is close to the second subglottal pole. Stevens (1998) suggests, however, that the second formant in proximity to the second subglottal pole is actually unstable. What is different in the two cases is that the subglottal pole is accompanied by a nearby zero, whereas formants are poles without accompanying zeros. Fant (1973) referred to the former as "bound poles" and to the latter as "free poles." Because the bound subglottal pole has an accompanying nearby zero, the potential stability of the region around it is

Lulich *et al.*: Subglottal resonances in speech perception

negated by the discontinuity between the pole and zero (cf. Fig. 1). Over a small span of frequencies around the center frequency of the pole-zero pair, the spectral effect changes quickly from being several dB negative to several dB positive. Since this span of frequencies is so small, a proposed stable target frequency for a formant near the pole would necessarily be small, contradicting the main force behind quantal theory.

V. CONCLUSION

The potential role of the second subglottal resonance in distinguishing front from back vowels has been noted before (Stevens, 1998), and there is evidence that speakers' vowel acoustic patterns are partially structured with reference to this subglottal resonance (Chi and Sonderegger, 2004; Lulich, 2006). We have shown that an acoustic discontinuity similar to that produced by the second subglottal resonance plays a role in the perception of non-steady-state vowels embedded in natural speech. This suggests that the discontinuity formed by the second subglottal resonance is actively used in the process of lexical access in speech perception.

ACKNOWLEDGMENTS

[1]We thank Xuemin Chi and Morgan Sonderegger for sharing these data with us from their study (Chi and Sonderegger, 2007).

[2]We thank Terrance Nearey (p.c.) for pointing out this alternative to us, as well as suggesting the response we have adopted.

Cheyne, H. A. (**2002**). "Estimating glottal voicing source characteristics by measuring and modeling the acceleration of the skin on the neck," Ph.D. thesis, MIT, Cambridge, MA.

Chi, X., and Sonderegger, M. (**2004**). "Subglottal coupling and vowel space," J. Acoust. Soc. Am. **115**, 2540–2540.

Chi, X., and Sonderegger, M. (**2007**). "Subglottal formats and vowel space,"

J. Acoust. Soc. Am. **122** (in press).

Chistovich, L. A. (**1985**). "Central auditory processing of peripheral vowel spectra," J. Acoust. Soc. Am. **77**, 789–805.

Cranen, B., and Boves, L. (**1987**). "On subglottal formant analysis," J. Acoust. Soc. Am. **81**, 734–746.

Fant, G. (**1973**). *Speech Sounds and Features* (MIT, Cambridge, MA).

Fant, G., Ishizaka, K., Lindqvist, J., and Sundberg, J. (**1972**). "Subglottal format," Speech Transmission Laboratory, Quarterly Status Progress Reports **1**, 1–12.

Flanagan, J., L. (**1957**). "Difference limen for formant amplitude," J. Speech Hear Disord. **22**, 205–212.

Grosjean, F. (**1996**). "Gating," Lang. Cognit. Processes **11**, 597–604.

Hanson, H. M. (**1996**). "Measurements of subglottal resonances and their influence on vowel spectra," J. Acoust. Soc. Am. **100**, 2656.

Hanson, H. M., and Stevens, K. N. (**1995**). "Sub-glottal resonances in female speakers and their effect on vowel spectra," Proceedings of the XIIIth International Congress of Phonetic Sciences **3**, 182–185.

Ishizaka, K., Matsudaira, M., and Kaneko, T. (**1976**). "Input acoustic-impedance measurement of the subglottal system," J. Acoust. Soc. Am. **60**, 190–197.

Ito, M., Tsuchida, J., and Yano, M. (**2001**). "On the effectiveness of whole spectral shape for vowel perception," J. Acoust. Soc. Am. **110**, 1141–1149.

Kiefte, M., and Kluender, K. R. (**2005**). "The relative importance of spectral tilt in monophthongs and diphthongs," J. Acoust. Soc. Am. **117**, 1395–1404.

Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857.

Lulich, S. M. (**2006**). "The role of lower airway resonances in defining vowel feature contrasts," Ph.D. thesis, MIT, Cambridge, MA.

Malme, C. I. (**1959**). "Detectability of small irregularities in a broadband noise spectrum," Research Laboratory of Electronics, Quarterly Progress Report **52**, 139–142.

Malme, C. I. (**1960**). "Detectability of small irregularities in a harmonic line spectrum," Research Laboratory of Electronics, Quarterly Progress Report **57**, 122–126.

Ross, B., and Pantev, C. (**2004**). "Auditory steady-state responses reveal amplitude modulation gap detection thresholds," J. Acoust. Soc. Am. **115**, 2193–2206.

Stevens, K. N. (**1972**). "The quantal nature of speech: Evidence from articulatory-acoustic data," in *Human Communication, A Unified View*, edited by J. E. E. David and P. B. Denes (McGraw-Hill, New York).

Stevens, K. N. (**1989**). "On the quantal nature of speech," J. Phonetics **17**, 3–45.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT, Cambridge, MA).

Stevens, K. N. (**2002**). "Toward a model for lexical access based on acoustic landmarks and distinctive features," J. Acoust. Soc. Am. **111**, 1872–1891.

Stevens, K. N., Kalikow, D. N., and Willemain, T. R. (**1975**). "A miniature accelerometer for detecting glottal waveforms and nasalization," J. Speech Hear Disord. **18**, 594–599.

Syrdal, A. K., and Gopal, H. S. (**1986**). "A perceptual model of vowel recognition based on the auditory representation of American English vowels," J. Acoust. Soc. Am. **79**, 1086–1100.

Dynamic spectral structure specifies vowels for children and adults[a)]

Susan Nittrouer[b)]

Speech and Hearing Science, Ohio State University, Columbus, Ohio 43210

When it comes to making decisions regarding vowel quality, adults seem to weight dynamic syllable structure more strongly than static structure, although disagreement exists over the nature of the most relevant kind of dynamic structure: spectral change intrinsic to the vowel or structure arising from movements between consonant and vowel constrictions. Results have been even less clear regarding the signal components children use in making vowel judgments. In this experiment, listeners of four different ages (adults, and 3-, 5-, and 7-year-old children) were asked to label stimuli that sounded either like steady-state vowels or like CVC syllables which sometimes had middle sections masked by coughs. Four vowel contrasts were used, crossed for type (front/back or closed/open) and consonant context (strongly or only slightly constraining of vowel tongue position). All listeners recognized vowel quality with high levels of accuracy in all conditions, but children were disproportionately hampered by strong coarticulatory effects when only steady-state formants were available. Results clarified past studies, showing that dynamic structure is critical to vowel perception for all aged listeners, but particularly for young children, and that it is the dynamic structure arising from vocal-tract movement between consonant and vowel constrictions that is most important. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2769624]

PACS number(s): 43.71.An, 43.71.Es, 43.71.Ft [PEI]   Pages: 2328–2339

I. INTRODUCTION

The question of how listeners assign vowel labels to acoustic signals has always been more problematic than the question of how consonant labels are assigned. Not long after experimenters began developing tools to independently manipulate the various acoustic properties of speech signals, the phenomenon of categorical perception was noted. This phenomenon was originally reported for experiments conducted in the 1950s involving consonant labeling where all acoustic properties were held constant in stimuli, except for one that was varied in discrete steps of equal size across a continuum (Liberman *et al.*, 1957). Results of such experiments are typically marked by flat labeling functions along the regions of the continua where stimuli are heard as belonging to a single phonemic category, and by sharp functions at category boundaries. Discrimination results show poor performance for stimuli in the flat, within-category regions, and contrastively good performance for stimuli straddling category boundaries. Experiments with vowels, however, failed to replicate the effects observed for consonants. Instead, labeling for vowel-like stimuli with formant frequencies varying along linear continua showed gradual slopes for the entire length of the continua, and discrimination among stimuli was good for all stimulus comparisons (Fry *et al.*, 1962). So it appeared in those early days of experimental speech perception that sensory processing for signal portions largely associated with vowels differed from processing for consonant-related signal portions.

Although there may have been no generally agreed upon perspective of *how* listeners process vocalic signals, the commonly accepted wisdom about *what* they process has always been that the frequencies of the first two or three steady-state formants are used by listeners in making vowel judgments (e.g., Ferrand, 2007). But this position faces challenges at even the most basic level: Continuous speech rarely has stretches that could be called steady state, and so the first challenge faced by the above-offered position has to do with the notion of "steady-state" syllable segments. Even if we modify the theoretical position to state that listeners make vowel judgments based on "target" formant frequencies, meaning the extremes in formant frequencies associated with specific vowels, other challenges arise. For example, target formant frequencies for the same vowel vary greatly across speakers due to differences in vocal-tract size and geometry: Not only are there length differences, but the ratio of oral cavity length to pharyngeal cavity length is very different for men, women, and children (e.g., Fant, 1973). Of course, problems of this sort have been handled with several proposals of how listeners might normalize across acoustic variation to derive stable phonetic representations. To deal with the problem of speaker variability, for example, various versions of speaker normalization have been offered (e.g., Gerstman, 1968; Syrdal and Gopal, 1986), and all share the theme that listeners must adjust their expectations of formant-to-vowel relations based on an individual speaker's acoustic characteristics, which become apparent by listening to that speaker. Although tests of speaker normalization have met with modest success (e.g., Ladefoged and Broadbent, 1957), these proposals alone cannot account for another problem, and that is the variability associated with phonetic

---
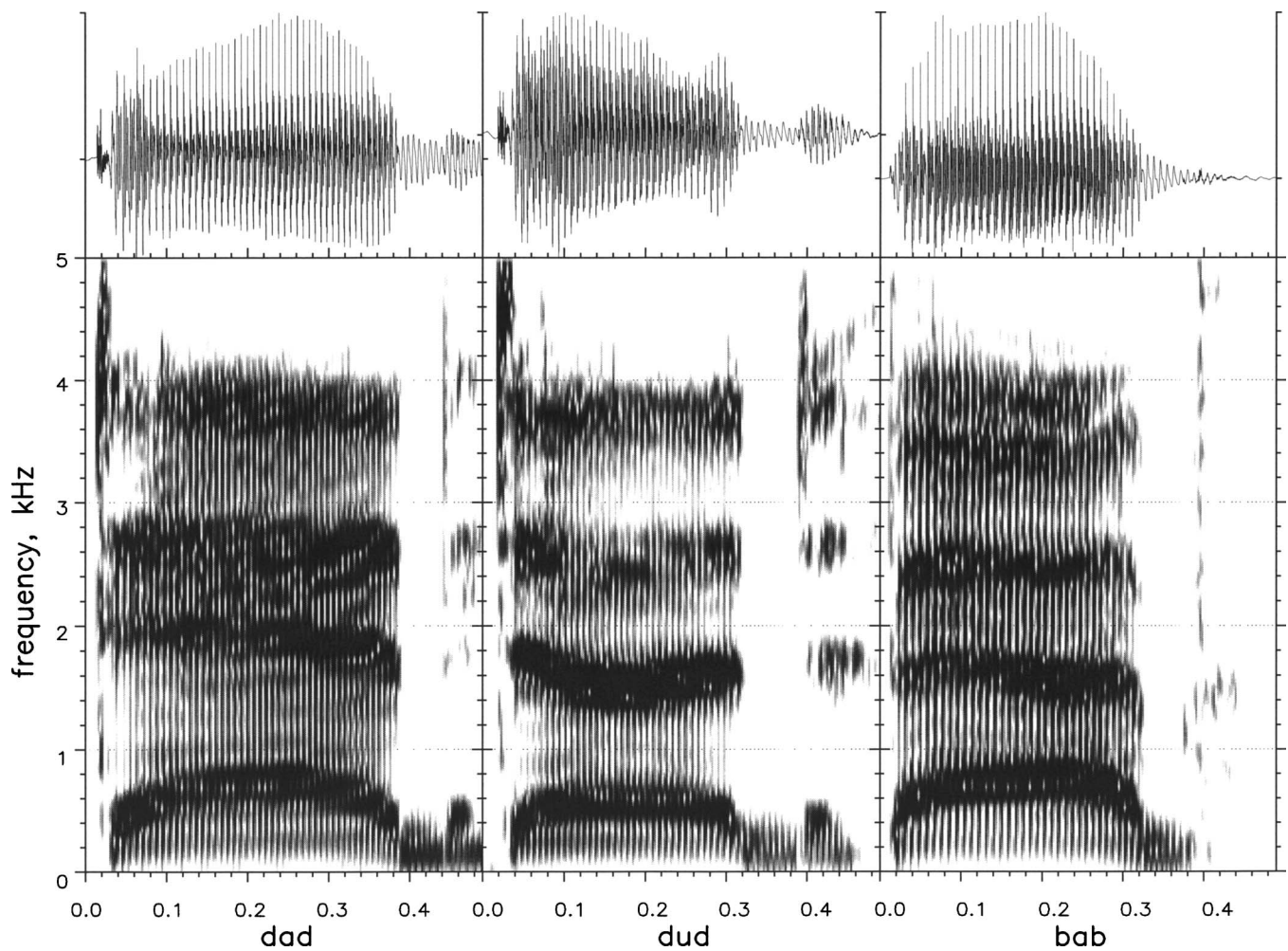
FIG. 1. Spectrograms of adult, male speaker saying "dad," "dud," and "bab."

context. This problem is illustrated in Fig. 1 showing spectrograms of "dad," "dud," and "bab," produced by an adult, male speaker. It can be seen that target F2 is more similar for "dud" and "bab" than for "dad" and "bab," even though "dad" and "bab" are heard as containing the same vowel. To handle the variability that arises due to phonetic context, the idea was proposed that listeners normalize their expectations of what formant frequencies should be for each vowel based on that context (e.g., Lindblöm and Studdert-Kennedy, 1967). So, if we combine notions of speaker and context normalization we must propose that listeners apply both kinds of normalization in the course of vowel perception. In this model, perception becomes complicated and might even be considered "mentalistic."

In contrast to these approaches, Strange and her colleagues explored the possibility that vowels are recognized not based on static formant frequencies recovered from specific regions of the signal, but rather based on patterns of change into and out of those specified signal regions. In a series of experiments these investigators replaced middle portions of natural consonant-vowel-consonant (CVC) syllables with equal amounts of silence (Jenkins et al., 1983; Strange et al., 1983, 1976). The syllables used in these experiments were obtained from several speakers, and typically nine different vowels were represented across the samples.

Listeners heard both the syllables with silent centers (i.e., the "vowel-less" syllables) as well as those extracted syllable centers presented in isolation. Results showed that listeners were more accurate in their vowel recognition for the vowel-less stimuli than for the steady-state formants. These findings led these investigators to conclude that "…dynamic acoustic information distributed over the temporal course of the syllable is utilized regularly by the listener to identify vowels" (Strange et al., 1976, p. 213). Specifically, Strange and her colleagues contended that it is the spectral change arising from the vocal tract moving from articulatory consonant configurations to vowel configurations, and back again, that specifies vowel identity. This conclusion has been supported by the work of others (e.g., Assmann and Katz, 2000; Fox, 1989), although some investigators have found that vowels can be labeled equally well with either static formants or dynamic formant transitions (e.g., Diehl et al., 1981).

Nearey and Assmann (1986) offered an alternative description of the notion that vowels are dynamically specified. They examined a variety of vowel productions, and observed that the syllabic regions specifically associated with vowel production have inherent spectral change that is uniquely associated with the vowel being produced. They further observed that these patterns of inherent spectral change are preserved in consonant contexts. From those observations,

Nearey and Assmann concluded that it is this inherent dynamic information that accounts for accurate vowel recognition, regardless of whether isolated vowels or vowels in consonant contexts are heard. Part of the support for their conclusion came from the finding that high error rates were observed when listeners heard only 30-ms portions of the signal region that could be termed the vowel target. Their tests showed that listeners needed stretches of vowel samples longer than 30 ms, presumably so they could hear the inherent spectral change. Nearey and Assmann did not present listeners with stimuli in which the middle portions were replaced by silence—the classic vowel-less stimuli. Nonetheless, the juxtaposition of their findings with those of Strange and colleagues offers two alternative views of dynamic vowel information: One view suggesting that it is precisely the movement between consonant syllable margins and vocalic targets that provides the relevant dynamic information, and the other suggesting that it is the production of the isolated vowel that provides that information. These alternatives influenced the design of stimuli used in this experiment.

But the above-described work was all done with adults as listeners. Murphy et al. (1989) were the first investigators to examine the perception of vowel-less syllables by children, using synthetic versions of /bæd/ and /bʌd/. One relevant finding from that study was that many children had difficulty labeling syllables when the deleted center portions were left silent. When those deleted portions were replaced with white noise, none of the children had difficulty labeling the vowels, prompting the conclusion that many children are unable to integrate two signal sections across a silent interval of even a few tens of milliseconds. Furthermore, the finding that all children performed accurately when listening to the syllables with filled centers led to the conclusion that children can use dynamic signal portions arising from the movement of the vocal tract between consonant- and vowel-related configurations for making vowel judgments, just as adults do.

Sussman (2001), on the other hand, used a different approach to investigate vowel perception by adults and children (4 to 5 years of age),[1] and obtained a somewhat different result. Stimuli were synthetic syllables designed to sound like /bib/ and /bæb/, with 40-ms transitions on either side of 280-ms regions of steady-state formants. Stimuli were manipulated in several ways. First, stimuli were presented as vowel-less stimuli. Children performed slightly (but not significantly) more poorly than adults in this condition, which would have been predicted based on the finding of Murphy et al. (1989) because there was nothing filling the silent centers. But it was another stimulus manipulation that was most important to Sussman's conclusions. In that manipulation, Sussman crossed the steady-state stimulus regions with the formant transitions at the syllable edges. In one condition the transitions and steady-state portions all supported the labeling of the same vowel (the *congruent* condition). In the other condition, 220-ms sections of the steady-state vocalic portions were removed and reinserted between the margins of the other stimuli (the *conflicting* condition). Results showed that all listeners performed with close to perfect accuracy in the *congruent* condition. When asked to label the stimuli in

the *conflicting* condition, all listeners responded with the label associated with the 220-ms steady-state section, but children did so slightly (but again not significantly) more often than adults. From this result, Sussman concluded that listeners of all ages, but especially children, weight steady-state formants most strongly in vowel recognition. Unfortunately, there are several reasons to worry that these stimuli may not inform us about how listeners recover vowel quality when hearing natural speech. In particular, common sense tells us that if a stimulus has a steady-state region as long as 220 ms then briefer stimulus portions, whether static or dynamic in nature, will have little chance of influencing decisions. And as Nearey and Assmannn's work (1986) emphasized, perfectly steady-state regions of this length are not found in natural speech, not even in natural vowels produced in isolation. Nonetheless, results of Sussman's experiment helped to renew interest in the question of what information child and adult listeners use in making vowel decisions.

The current study extended the work of Murphy et al. (1989) and Sussman (2001) in order to examine the patterns of vowel perception by children. In particular, this study was designed to cross the kind of vowel contrast being examined (i.e., vowel height or vowel frontedness) with the expected magnitude of influence on vowel formant frequencies, as well as on transitions, introduced by the consonant context (i.e., expected to be highly influential or not very influential). A strong motivation behind the design of stimuli in this experiment was the lack of articulatory complexity in the stimuli used by Murphy et al. and Sussman, and so a lack of opportunity for strong coarticulatory effects. Murphy et al. had examined a contrast between two vowels that differed primarily in tongue frontedness (/æ/vs /ʌ/), whereas Sussman examined a contrast between two vowels that primarily involved jaw height (/i/ vs /æ/). Both used a bilabial consonant (/b/), although Murphy et al. also used an alveolar at the final syllable margin. Because the lips can move largely independently of the jaw and tongue, little coarticulatory effect would be predicted for the /b/ consonantal frame. For the present study it was hypothesized that canonical vowel formant frequencies might be preserved more or less robustly depending on the type of contrast (height versus frontedness), and that different consonant contexts should affect those frequencies, as well as formant transitions, to different extents. Specifically, the hypothesis was that the perception of vowel contrasts differing in whether the tongue body is more fronted or backed (such as /æ/ vs /ʌ/) should be more influenced by the type of consonant context (whether it involves the lips or tongue tip) than the perception of vowel contrasts that differ in jaw height (such as /i/ vs /æ/) because the tongue body is tightly linked to tongue tip movement whereas the jaw is not. In other words, consonant contexts that constrain tongue movement (such as /dVd/) should influence decisions about vowels differing in frontedness more than consonant contexts that do not involve the tongue (such as /bVb/). These potential relations of contrast type and consonant context were considered in stimulus design for this experiment.

Stimuli were created to be as natural as possible in this experiment. To achieve this goal, natural productions from

adult speakers were used. Because Murphy *et al.* (1989) had observed that some children had difficulty integrating syllable portions across a silent gap, deleted syllable centers were replaced with coughs in the vowel-less stimuli. Coughs were considered more ecologically valid than white noise, and using them replicated the practice of Warren (1970; Warren and Obusek, 1971), who used coughs to replace syllable portions in the study of phoneme restoration. Perceptual impressions of such stimuli typically are that someone is saying a word, and someone else coughs at the same time, partially masking the word production. That is, the cough streams off from the string of words being heard. This impression is well-documented in the work of Warren, and was obtained in the current study as well. Results of vowel decisions for these stimuli were compared to decisions for steady-state vowel formants. Children, as well as adults, served as listeners because we were explicitly interested in how children label vowels, and if they differ from adults in how they do so.

A methodological problem that needs to be anticipated in experiments involving vowel-less stimuli is the possibility that vowel labeling may be highly accurate overall. For example, Strange *et al.* (1976) reported that listeners correctly recognized vowels 90.5% of the time when tokens from a single talker (producing nine vowels) were presented. Murphy *et al.* (1989) reported correct recognition of better than 90% for both adult and child listeners when vowel decisions were binary, and Sussman (2001) obtained recognition scores of close to 90%. Overall rates of responding can be diminished by using tokens from many different speakers (more than the three used in this experiment) or by forcing listeners to select from more than two vowel choices. On the other hand, if great uncertainty is incorporated into stimulus design and results show age-related differences it is hard to determine if the source of those differences is perception itself, or the fact that children have difficulty listening under conditions of high uncertainty (e.g., Wightman and Kistler, 2005). Because of that concern about eventual interpretation, stimulus uncertainty was restricted in this experiment, even though stimuli were designed to maximize coarticulatory effects. There was no way to know ahead of time if overall performance would be high, as in earlier experiments, or if it would be diminished substantially in this experiment precisely because stimulus design was meant to maximize coarticulatory effects.

II. METHOD

A. Listeners

Forty eight adults, 55 seven year olds, 56 five year olds, and 62 three year olds came to the laboratory to participate. All adults were between 20 and 40 years of age. All 5 and 7 year olds were between −1 and +5 months of their birthdays. Three year olds were between 3 years, 5 months, and 3 years, 11 months.[2] No more than a 40%-to-60% split between male and female participants in each age group was permitted. All participants were native English speakers with no histories of speech, language, or hearing problems. Children were required to have had fewer than six episodes of

TABLE I. The number of participants of each age, in each contrast.

	/bɪb/vs/bæb/	/bæb/vs/bʌb/	/dɪd/vs/dæd/	/dæd/vs/dʌd/
3 year olds	9	12	11	10
5 year olds	14	13	14	15
7 year olds	13	12	12	18
Adults	12	12	12	12

otitis media before their third birthdays. All participants were required to pass hearing screenings of the frequencies 0.5, 1, 2, 4, and 6 kHz presented at 25 dB HL to each ear separately. Children needed to perform at or better than the 30th percentile on the Goldman-Fristoe Test of Articulation 2, Sounds-in-Words subtest (Goldman and Fristoe, 2000), and adults needed to read at or better than an 11th grade reading level on the Wide Range Achievement Test—Revised (Jastak and Wilkinson, 1984). Although superficial, this brief screening provided some evidence that our adults had normal language. We took the fact that none of the children participating in the experiment were being seen for language or reading intervention as evidence that they had typical language for their ages. Our screening procedures resulted in some attrition for 3 year olds: 8 three year olds simply refused to cooperate with even the screening procedures, seven failed the hearing screening, and five failed the Goldman-Fristoe Test of Articulation. Consequently, only 42 three year olds were left to participate.

Different listeners were assigned to participate with each of the four contrasts: (1) /bɪb/ vs /bæb/; (2) /bæb/ vs /bʌb/; (3) /dɪd/ vs /dæd/; and (4) /dæd/ vs /dʌd/. Table I shows how many participants in each age group participated in testing with each contrast. Numbers for 3 year olds do not sum to 62 because some 3 year olds did not pass the pretest (reported in Sec. III).

B. Stimuli

Natural stimuli were used in this experiment, and presented both with and without consonantal context. Three vowels were used: /ɪ/, /æ/, and /ʌ/. This is fewer than have been used in vowel labeling experiments with adults, but more than were used with children in the experiments of Murphy *et al.* (1989) or Sussman (2001). The choice of these three vowels meant that they could be paired so that listeners were being asked to make a decision about vowel height (/ɪ/ vs /æ/) or a decision about frontedness (/æ/ vs /ʌ/). For this experiment, the consonant contexts of /dVd/ and /bVb/ were used. These two contexts were selected to vary the extent of contextual influence on vowel formant frequencies. This can be seen in Fig. 1, where /ʌ/ F2 frequency in /dʌd/ and /æ/ F2 frequency in /bæb/ are similar at the middle of the syllables, but the direction of transitions into and out of those vowel targets differ for the two syllables.

Stimuli differed from those of Murphy *et al.* and Sussman in several ways. The stimuli in both of those earlier experiments were synthetic; these stimuli were natural. Murphy *et al.* used only a /æ/ vs /ʌ / contrast and Sussman used only a /æ/ vs /i/ contrast. In this experiment, both vowel contrasts were examined, with /ɪ/ substituted for the /i/ of

TABLE II. Formant frequencies for syllables produced by three speakers, two tokens per speaker.

	F1			F2			F3		
	Onset	Middle	Offset	Onset	Middle	Offset	Onset	Middle	Offset
/bæb/	449	721	567	1575	1694	1339	2360	2498	2437
	(58)	(26)	(100)	(68)	(132)	(80)	(76)	(143)	(149)
/bɪb/	370	438	424	1762	1830	1518	2512	2596	2476
	(9)	(11)	(57)	(46)	(71)	(52)	(106)	(140)	(118)
/bʌb/	463	624	510	1217	1278	1199	2505	2633	2559
	(50)	(14)	(50)	(47)	(45)	(35)	(121)	(104)	(91)
/dæd/	388	704	488	1852	1708	1612	2649	2494	2616
	(41)	(49)	(47)	(65)	(150)	(69)	(174)	(141)	(217)
/dɪd/	327	417	350	1913	1967	1784	2681	2624	2641
	(22)	(42)	(77)	(141)	(134)	(98)	(146)	(115)	(128)
/dʌd/	395	599	427	1712	1461	1633	2584	2588	2667
	(37)	(25)	(42)	(109)	(108)	(54)	(177)	(159)	(112)

Sussman's experiment to make the difference in formant frequencies between vowels in each contrast more similar. The stimuli in those earlier experiments did not include a condition with a /dVd/ context, which provides tongue constraints on either side of the vowel target: Murphy *et al.* used a /bVd/ context and Sussman used a /bVb/ context. Because /b/ involves lip and jaw gestures, it would not be expected to affect vocalic F2 very much, if at all. In the current experiment, the /dæd/ vs /dʌd/ contrast was the one expected to have vocalic formant frequencies affected most strongly by consonantal context because the /æ/ vs /ʌ/ contrast largely involves a distinction in the fronting of the tongue body, and the production of /d/ involves the tongue. In fact, this prediction is readily observed in the F2 frequencies reported in Tables II and III: For the /æ/ and /ʌ/ vowels produced either in isolation or in the /bVb/ context, the differences in F2 are roughly 400 Hz. However, when these vowels were produced in the /dVd/ context that difference is reduced to roughly 240 Hz, mainly because F2 in /dʌd/ is raised considerably, reflecting tongue fronting. Consequently, we would expect listeners to have a particularly rough time labeling vowels based on target formant frequencies, in the absence of consonantal context, for this particular contrast.

Vowel and syllable samples were recorded by three adult, male speakers who had fundamental frequencies (f0s) of roughly 130 Hz, with no extreme breathiness or fry. Each speaker was recorded producing five samples each of /bɪb/, /bæb/, /bʌb/, /dɪd/, /dæd/, /dʌd/, /ɪ/, /æ/, and /ʌ/. First, samples of the six syllables were obtained in five randomized sets, and then samples of the three isolated vowels were obtained

TABLE III. Formant frequencies for isolated vowels produced by three speakers, two tokens per speaker.

	F1	F2	F3
/æ/	786	1694	2524
	(26)	(98)	(168)
/ɪ/	452	1963	2627
	(14)	(113)	(77)
/ʌ/	671	1314	2487
	(16)	(64)	(106)

in five randomized sets. Speakers were instructed to produce all syllables as similar in duration as possible, and to produce all isolated vowels as similar in duration as possible. These instructions resulted in syllables with roughly 300 ms of vocalic portion and isolated vowels of roughly 200 ms. Speakers were also instructed to produce samples with similar inflectional patterns and similar f0. Samples were digitized as they were produced at a 22.05-kHz sampling rate with 16-bit digitization. If a sample was produced with vocal fry, with f0 extremely different from 130 Hz, or with duration extremely different from the targets of 300 ms (for vocalic portions of syllables) and 200 ms (for isolated vowels), the speaker was asked to repeat the production. All productions were analyzed to derive whole syllable or vowel duration, duration of vocalic portion only (for syllables), f0, and frequencies of the first three formants at vocalic onset, vocalic middle, and vocalic offset (i.e., just before closure). The two productions of each type from each speaker that were most representative of that production type from that speaker were used as stimuli in the listening experiment. Table II displays mean F1, F2, and F3 frequencies at the onset of the vocalic portion, vocalic middle, and the offset of the vocalic portion for the syllables, and Table III shows mean formant frequencies for vowels spoken in isolation, at the middle of the sample. Although the temporal middle of a speech sample does not always line up perfectly with the "target" (i.e., most extreme) formant frequencies, it is the best single time frame that provides a close estimate of those frequencies for all formants. A separate speaker, a woman, was recorded producing several coughs, and sections of these coughs were used in stimulus creation as well. A cough from a female speaker helped to ensure the perceptual impression that whole syllables were being heard, with someone else coughing during their production. Six kinds of stimuli were created, and are described in the following section. rms amplitude of all stimuli was equalized.

Flat stimuli: Three kinds of 150-ms stimuli were created to sound like isolated vowel productions. (1) 150-ms stretches of the vowels produced in isolation were obtained by finding the temporal middle of the productions, and taking 75 ms on either side of that (the "isolated vowel"
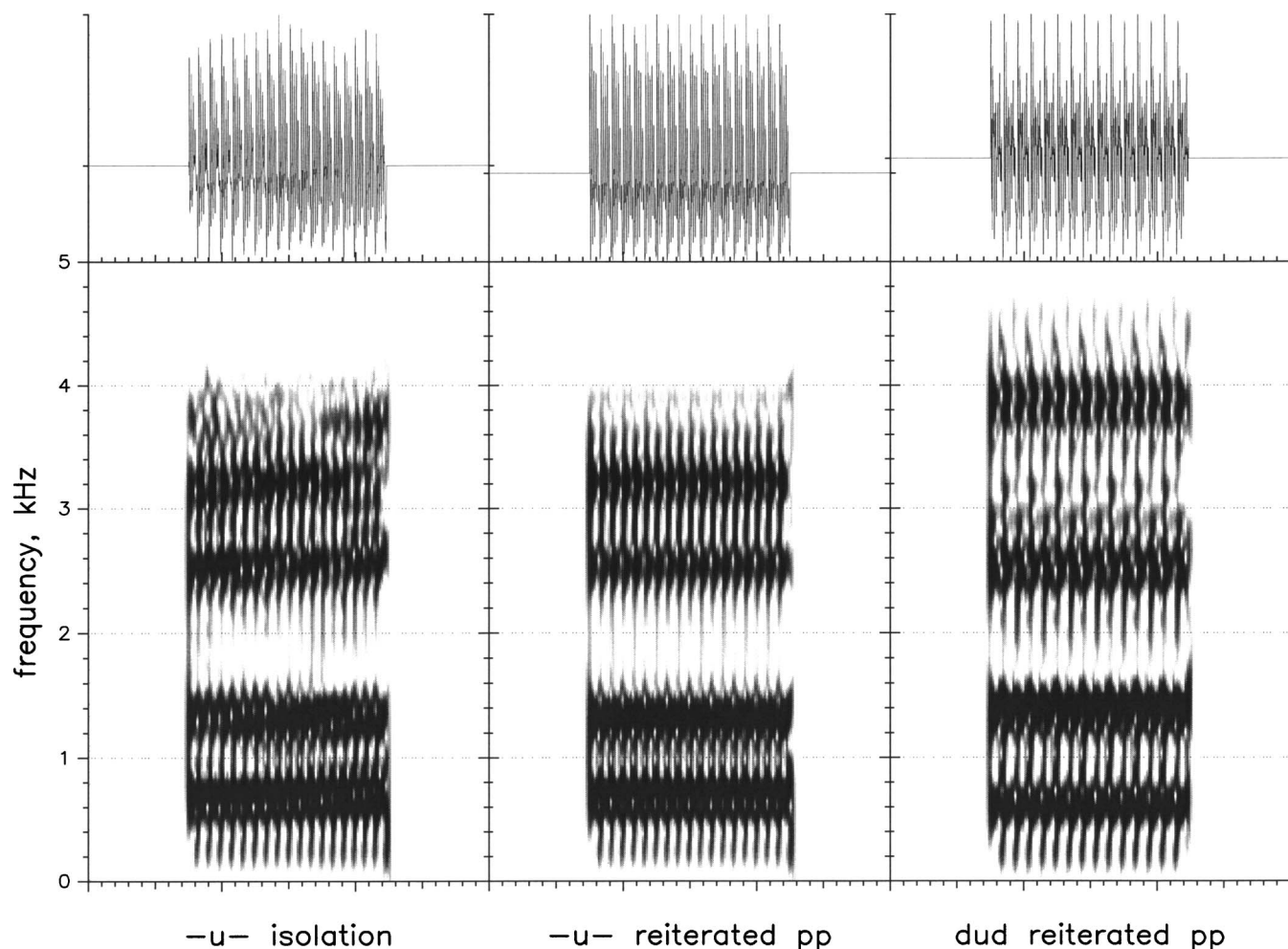
FIG. 2. Spectrograms of the three examples of the flat stimuli: 150 ms of the vowel /ʌ/ produced in isolation (labeled here as -u-); reiterated pitch periods from the vowel /ʌ/ produced in isolation; and reiterated pitch periods from /ʌ/ produced in the syllable "dud."

stimuli); (2) the two pitch periods closest to the temporal middle of each isolated vowel were reiterated ten times to produce 150-ms stretches of steady-state formants (the "reiterated vowel" stimuli); and (3) the two pitch periods closest to the temporal middle of the vocalic portions of the syllables were reiterated ten times to produce 150-ms stretches of steady-state formants (the "reiterated syllable" stimuli). These last two types of stimuli were created to address Nearey and Assmann's (1986) hypothesis that vowel-inherent dynamic spectral change supports vowel labeling more strongly than steady-state information: The isolated vowel productions exhibited inherent spectral change; the reiterated pitch periods did not. The inclusion of reiterated pitch periods from the isolated vowel and from syllable centers meant that some stimuli had formant frequencies consistent with isolated vowel production, and some had formant frequencies consistent with those found in coarticulated syllables. All waveform editing was done with cuts made at zero crossings. Figure 2 shows examples of the three kinds of flat stimuli: a section of a vowel produced in isolation, reiterated pitch periods from that vowel, and reiterated pitch periods from a syllable with that vowel as the nucleus.

Dynamic stimuli: (1) Whole, unaltered syllables were used (the "whole syllable" stimuli); (2) the middle 50% of

the syllable was replaced with a cough section (the "50% cough" stimuli); and (3) all of the syllable except for the first and last three pitch periods was replaced with a cough section (the "pitch period" stimuli). For the 50%-cough stimuli, 150-ms stretches consisted of cough, on average. There were generally 10 pitch periods preceding that stretch of cough (75 ms) and 10 pitch periods following that stretch of cough (again, roughly 75 ms). For the pitch-period stimuli, there was generally 255 ms of cough separating the three pitch periods on either side. Thus, about 85% of the syllable was replaced with cough. Figure 3 shows examples of the three kinds of dynamic stimuli: a whole, unaltered syllable, the same syllable with the center 50% replaced by the cough, and the same syllable with all but the first and last three pitch periods replaced by the cough.

C. Equipment and materials

All speech samples were recorded in a sound-proof booth, directly onto the computer hard drive, via an AKG C535 EB microphone, a Shure M268 amplifier, and a Creative Labs Soundblaster 16-bit analog-to-digital converter. A waveform editor (WAVED; Neely and Peters, 1992b) was used for recording and editing. An acoustic analysis program based in MATLAB was used for spectral analysis, and
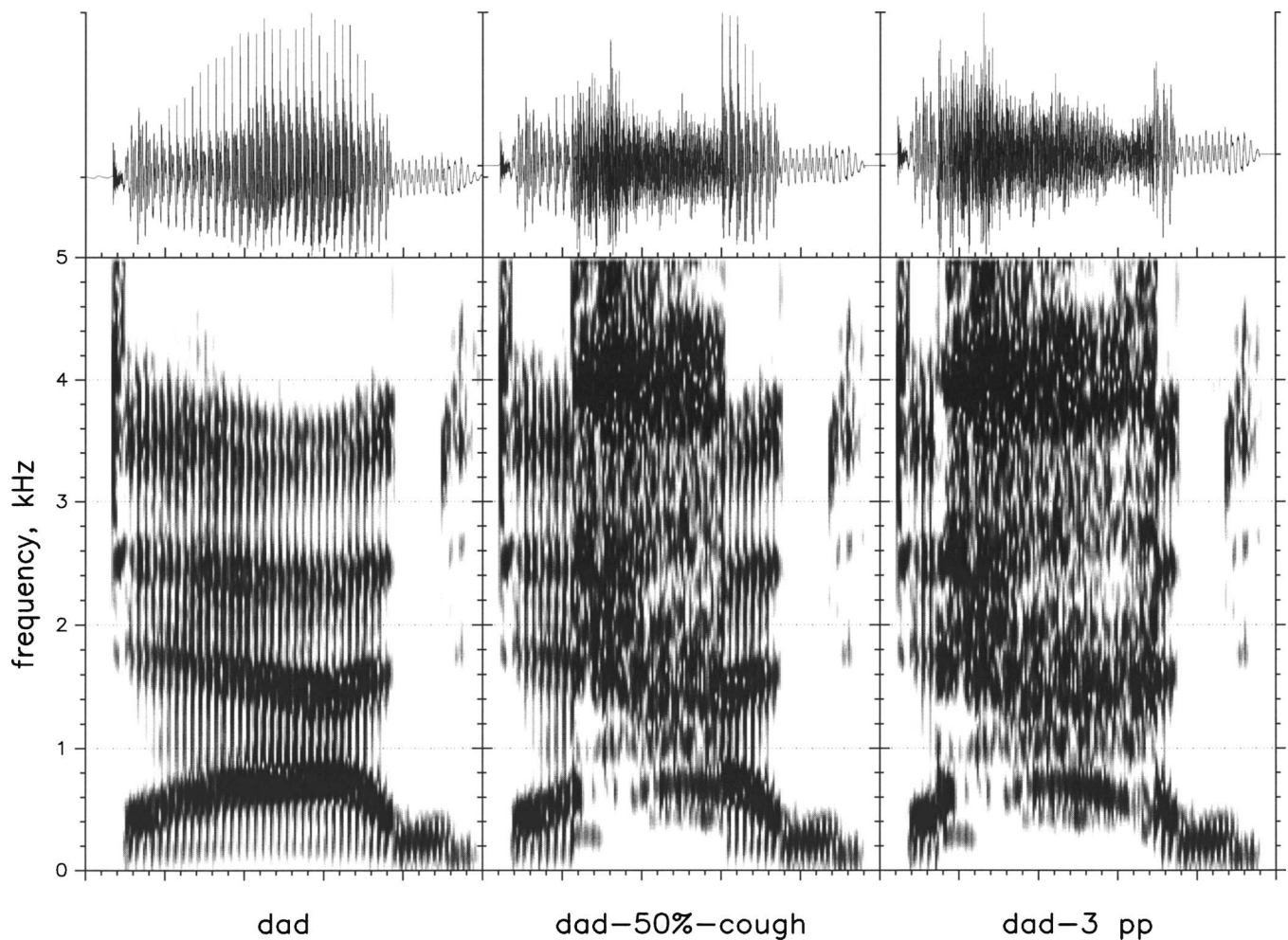
FIG. 3. Spectrograms of three examples of the dynamic stimuli: the whole syllable "dad;" the same syllable with the center 50% replaced by a cough; and the same syllable with all except the first and last three pitch periods replaced by a cough.

SPECTO (Neely and Peters, 1992a) was used to make spectrograms.

Perceptual testing took place in a sound-proof booth, with the computer that controlled the experiment in an adjacent room. The hearing screening was done with a Welch Allen TM262 audiometer and TDH-39 earphones. Stimuli were stored on a computer and presented through a Creative Labs Soundblaster card, a Samson headphone amplifier, and AKG-K141 headphones. The experimenter recorded responses with a keyboard connected to the computer. Two hand-drawn pictures (8 in. × 8 in.) were used to represent each response label in each experiment. These included a baby's bib for *bib*, bubbles for *bub*, a baby babbling for *bab*, a child showing his mother a picture that he made (and supposedly saying "Look what I did.") for *did*, a man for *dad*, and a firecracker with the fuse burnt for *dud*. Game boards with ten steps were also used with children: They moved a marker to the next number on the board after each block of test stimuli. Cartoon pictures were used as reinforcement that the child had completed another block of stimuli, and these pictures were presented on a color monitor after completion of each block of stimuli. A bell sounded while the pictures were being shown and served as additional reinforcement.

D. Procedures

Listeners came to the laboratory for one session. Every participant was first given the hearing screening, followed by either the speech screening (for children) or the reading screening (for adults). Next, the flat and dynamic stimuli were presented, in randomized orders across the participants. Presentation with the flat stimuli included the 150-ms segments of isolated vowel, the reiterated pitch periods from the isolated vowel, and the reiterated pitch periods from the syllable center. Presentation with the dynamic stimuli included the whole syllables, the 50% cough syllables, and the pitch period syllables. Presentation with both kinds of stimuli consisted of ten blocks of all 36 stimuli (3 stimulus types × 2 vowels × 3 speakers × 2 tokens).

Before testing with each stimulus type, participants were presented with either the isolated vowels or the whole syllables, depending on stimulus type to be presented, and asked to match them to the pictures used to represent them. All listeners, regardless of age, were given the same instructions. For practice with the isolated vowels, listeners were told that they would not be hearing all of the word, but they should still point to the correct picture. The pictures were

TABLE IV. Mean percent correct vowel recognition for each contrast, across all listeners and syllable types.

/bɪb/vs/bæb/	/bæb/vs/bʌb/	/dɪd/vs/dæd/	/dæd/vs/dʌd/
97	97	94	89
(6)	(6)	(3)	(4)

TABLE VI. Mean percent correct vowel recognition for each syllable type, across all contrasts and listeners.

Flat			Dynamic		
Isolated vowel	Reiterated vowel	Reiterated syllable	Whole syllable	50% cough	3 pitch periods
98	98	96	99	96	80
(7)	(5)	(7)	(2)	(7)	(18)

introduced one at a time by the experimenter, and the appropriate label told to the listener. Generally there was a brief discussion about how the picture related to the word (e.g., *bub* is a little bit of *bubble*), and listeners were given practice responding to live voice. Listeners were asked to point to the appropriate picture, and say the word they heard, as the form of responding. Having both kinds of response helped to ensure that the listener was matching what was heard to the corresponding picture. Next, listeners heard all 12 natural, unedited stimuli, either isolated vowels or whole syllables (2 vowels × 3 speakers × 2 tokens), and had to respond to 11 of the 12 stimuli correctly to proceed to testing. During this practice activity, the experimenter listened to the stimuli under headphones. If a child had some initial difficulty matching pictures to stimuli, feedback was provided for up to two blocks of the 12 stimuli. By the third block of stimuli the child had to be able to respond with no feedback. If the listener could not do so, that listener was not tested in that condition (either flat or dynamic stimuli). This pretest served as a check that all listeners were able to match words (or word portions) to pictures. When testing started, listeners were told to continue doing the same thing that they had been. No more feedback was provided at this point, and the experimenter removed her headphones so that she did not know which stimulus was being presented on any given trial.[3]

It is well-documented that children have difficulty recognizing single phonemes (e.g., Routh and Fox, 1984; Liberman *et al.*, 1974; Walley *et al.*, 1986), and so it may be that they have difficulty following instructions to label them. So, in this experiment, listeners were not asked to label individual vowels, but rather were asked to decide between words containing those vowels. Moreover, uncertainty escalates as choices increase, so only two words can be used in each task with children. These considerations contributed to decisions regarding how to modify the experimental task from that used with adults.

The dependent measure in this experiment was the percentage of stimuli in each condition recognized as containing the originally produced vowel. Statistical analyses were performed on percent correct scores to examine patterns across contrasts, listener age, and stimulus types. Arcsine transforms were used because percentages were generally close to 100% correct, and so were highly skewed and kurtotic. A

TABLE V. Mean percent correct vowel recognition for each age group, across all contrasts and syllable types.

Adults	7 year olds	5 year olds	3 year olds
98	95	93	90
(2)	(4)	(6)	(9)

screening of the transformed data showed that this transformation mitigated the effects of skewness and kurtosis sufficiently to allow analyses of variance (ANOVAs) to be performed.

III. RESULTS

Some of the 3 year olds had difficulty reaching the criterion for participation in the pretest. For the /bɪb/vs /bæb/ contrast, 3 three year olds could not reach this criterion for isolated vowels, but all reached criterion for whole syllables. For the /bæb/ vs /bʌb/ contrast, all 3 year olds reached criterion for isolated vowels, but two did not do so for whole syllables. For the /dɪd/ vs /dæd/ contrast, 2 three year olds could not reach the oriterion for isolated vowels, and a different 2 three year olds could not reach criterion for whole syllables. For the /dæd/ vs /dʌd/ contrast, only 1 three year old could not reach criterion for isolated vowels, but all reached criterion for whole syllables. None of the older children or adults had difficulty reaching the criterion to participate in the actual testing for any of the contrasts.

Tables IV–VI show percent correct responses (and standard deviations) for each contrast, age group, and stimulus type; the Appendix shows means for each group, in each condition. Three general trends are apparent: (1) /dæd/ vs /dʌd/ showed less accurate recognition than other contrasts; (2) the pitch period stimuli showed less accurate recognition than other types of stimuli; and (3) younger listeners were less accurate than older listeners. A three-way ANOVA confirmed that these three factors (contrast, stimulus type, and age) did indeed show significant effects for percent correct vowel recognition: contrast, $F(3,173)=62.34$, $p<0.001$; stimulus type $F(5,865)=266.84$, $p<0.001$; and age, $F(3,173)=42.86$, $p<0.001$. However, there were also significant interactions: Contrast × Stimulus type, $F(15,865)=26.13$, $p<0.001$; Contrast × Age, $F(9,173)=2.49$, $p=0.011$; Stimulus type × Age, $F(15,865)=8.18$, $p<0.001$; and the three-way interaction of Contrast × stimulus type × Age, $F(45,865)=1.57$, $p=0.011$. Consequently, these results do nothing to identify how children's performance varied across stimulus type and contrast, or how it differed from that of adults. To look at these questions, results for each stimulus type were examined separately to ascertain the effects of contrast and listener age on performance, and to see where interactions occurred. These analyses were critical to determining if the extent of coarticulation affected vowel labeling.

Table VII shows results of two-way ANOVAs performed on data for each stimulus type separately, with contrast and age as the main effects. For each stimulus type there were

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Susan Nittrouer: Dynamic spectral structure of vowels   2335

TABLE VII. Results of a two-way ANOVA for all four contrasts, for each stimulus type separately. Denominator degrees of freedom are shown in the first line of each section, and reflect the fact that both contrast and age are between subjects factors. Numerator degrees of freedom are shown below. Exact *p* values are shown if less than 0.05. NS indicates nonsignificant effects.

Stimulus type	df	F	p
Flat stimuli			
Isolated vowel	177		
Contrast	3	15.04	<0.001
Age	3	19.74	<0.001
Contrast×Age	9	2.44	0.012
Reiterated vowel	177		
Contrast	3	3.86	0.011
Age	3	18.20	<0.001
Contrast×Age	9	2.32	0.017
Reiterated syllable	177		
Contrast	3	26.45	<0.001
Age	3	15.40	<0.001
Contrast×Age	9	2.07	0.035
Dynamic stimuli			
Whole syllable	181		
Contrast	3	3.50	0.017
Age	3	12.48	<0.001
Contrast×Age	9	0.47	NS
50% cough	181		
Contrast	3	16.80	<0.001
Age	3	19.36	<0.001
Contrast×Age	9	0.86	NS
Pitch periods	181		
Contrast	3	108.87	<0.001
Age	3	38.64	<0.001
Contrast×Age	9	1.75	NS

TABLE VIII. Cohen's *d* for mean percent correct vowel recognition in the /bɪb/ vs /bæb/ and the /dæd/ vs /dʌd/ contrasts for the reiterated syllable and 50% cough stimuli. This metric provides an index of the magnitude of coarticulatory effect on listeners' responses.

	Reiterated syllable	50% cough
3 year olds	1.39	0.84
5 year olds	1.25	0.93
7 year olds	1.18	0.62
Adults	0.61	1.00

significant effects of contrast and age, indicating that all groups generally performed more poorly for the /dæd/ vs /dʌd/ contrast than for other contrasts and that younger listeners generally performed more poorly than older listeners. Of greater relevance for this study, there were significant Contrast×Age interactions for flat stimuli, but not for dynamic stimuli. In other words, children's perception (compared to adults' perception) was disproportionately affected by the contrast only when stimuli were flat. When stimuli were dynamic, all listeners were similarly affected by the extent of coarticulatory effects in those stimuli. Children performed disproportionately more poorly than adults for flat stimuli involving the /dæd/ vs /dʌd/ contrast than for the other contrasts. This is precisely the contrast in which are found the greatest coarticulatory constraints of the consonantal context on target formant frequencies for the vowel. It is tempting to hypothesize that this result indicates that young children's perception is disrupted in the face of heavy coarticulatory effects. That may indeed be true, as we find here that young children's labeling of flat stimuli was generally diminished for the /dæd/ vs /dʌd/ contrast compared to conditions in which coarticulatory effects were not as strong. One piece of evidence for this assertion is the finding that 3-year-olds' labeling of isolated vowel productions of /æ/ and /ʌ/ was poorer when those same stimuli were presented along with reiterated pitch periods from the syllables /dæd/

and /dʌd/ than when they were presented along with reiterated pitch periods from /bæb/ and /bʌb/: the Appendix shows that mean recognition of these isolated vowels was 95.67 percent correct in the /bæb/ vs /bʌb/ condition, but only 83.44 percent correct in the /dæd/ vs /dʌd/ condition. So, young children encountered difficulty labeling any stimuli heard as isolated vowels when some of those stimuli were reiterated pitch periods from heavily coarticulated syllables.

A metric used to examine the apparent finding that young children have disproportionate difficulty labeling vowels from steady-state formants when coarticulatory effects are strong, as in the /dæd/ vs /dʌd/ contrast, was Cohen's *d*. This statistic is the difference between any two means, normalized by the pooled standard deviation associated with those means (Cohen, 1988). While straightforward in computation, Cohen's *d* serves as a robust index of effect size. In this case, it can index the magnitude of coarticulatory effects on vowel labeling. To this end, means of percent correct vowel recognition were compared between a contrast with little expectation of coarticulatory effects (/bɪb/ vs /bæb/) and the contrast with the greatest expectation of coarticulatory effects (/dæd/ vs /dʌd/). Two stimulus types were examined: a set of flat stimuli, the reiterated syllable stimuli, and a set of dynamic stimuli, the 50% cough stimuli. These stimulus types were selected because they best represent what listeners are likely to hear in the real world: We normally hear whole words rather than isolated vowels, and so any acoustic consequences of coarticulation are available, even if some syllable portions are masked. Table VIII shows Cohen's *d* for each age group, for each of these syllable types, computed as the difference between mean percent correct scores for the /bɪb/ vs /bæb/ and /dæd/ vs /dʌd/ contrasts, divided by the pooled deviations of those means. What we find is that all three groups of children showed greater coarticulatory effects on their vowel labeling when only steady-state formant frequencies were presented (i.e., reiterated syllable stimuli) than when dynamic syllable components were presented (i.e., 50% cough stimuli). This clearly was not the case for adults, whose performance held up well with the reiterated syllable stimuli derived from /dæd/ and /dʌd/.

Another useful way of viewing these data is to examine trends across syllable types for each contrast, for each age group separately. These analyses should provide information about how listeners of each age group dealt with the degraded stimuli compared to the unprocessed stimuli: Were different age groups more or less affected either by having truly steady-state formants (as in the reiterated conditions) rather than isolated vowels (with inherent spectral change) or

by having coughs replace vocalic centers? To help answer these questions, a series of matched *t*-tests were conducted for each listener age, for each contrast. These *t*-tests were computed for the following comparisons: (1) isolated vowel versus reiterated vowel; (2) isolated vowel versus reiterated syllable; (3) whole syllable versus 50% cough; (4) whole syllable versus pitch period; (5) isolated vowel versus whole syllable; and (6) reiterated syllable versus 50% cough. The last comparison was considered critical because it involved the acoustic portions that children are likely to hear in their daily lives. Again, arcsine transformations were used, and an alpha level of 0.01 was applied, rather than 0.05, due to the high probability of obtaining significant differences by chance when so many comparisons are made. A Bonferroni correction may also be applied, and with six contrasts per condition this would mean that significance at the 0.01 level would be reached only when obtained *p* values are less than 0.002. Results reaching this significance level are noted by an asterisk.

/bɪb/ vs /bæb/: All three children's groups performed more poorly for the pitch-period stimuli than for the whole syllables: 3 year olds, $t(8)=7.08$, $p<0.001^*$; 5 year olds, $t(13)=7.42$, $p<0.001^*$; and 7 year olds, $t(12)=3.75$, $p=0.003$. No other differences were observed.

/bæb/ vs /bʌb/: All four age groups showed poorer performance for the pitch-period stimuli than for the whole syllables: 3 year olds, $t(9)=7.43$, $p<0.001^*$; 5 year olds, $t(12)=10.02$, $p<0.001^*$; 7 year olds, $t(11)=4.77$, $p<0.001^*$; and adults, $t(11)=3.25$, $p=0.008$. In addition, 5 year olds performed more poorly for isolated vowels than for whole syllables, $t(12)=3.78$, $p=0.003$. The Appendix shows that 3 year olds performed similarly to 5 year olds with both isolated vowels and whole syllables: The mean difference between the two conditions was 3 percentage points for both groups. However, this comparison for 3 year olds did not quite reach criterion for reporting statistical significance because of the fewer degrees of freedom.

/dɪd/ vs /dæd/: Again, all four age groups showed poorer performance for the pitch-period stimuli than for the whole syllables: 3 year olds, $t(8)=6.12$, $p<0.001^*$; 5 year olds, $t(13)=9.02$, $p<0.001^*$; 7 year olds, $t(11)=6.64$, $p<0.001^*$; and adults, $t(11)=5.60$, $p<0.001^*$. In addition, 5 year olds performed more poorly for the 50% cough stimuli than for whole syllables, $t(13)=3.25$, $p=0.006$. In this case, the failure to find a similar result for 3 year olds may be attributable both to fewer degrees of freedom and to the fact that there was not as great of a difference between the two conditions for 3 year olds: mean differences between the two conditions were 3 and 8 percentage points for 3 and 5 year olds, respectively.

/dæd/ vs /dʌd/: Once again, all four age groups showed poorer performance for the pitch-period stimuli than for the whole syllables: 3 year olds, $t(9)=15.14$, $p<0.001^*$; 5 year olds, $t(14)=25.81$, $p<0.001^*$; 7 year olds, $t(17)=22.46$, $p<0.001^*$; and adults, $t(11)=17.58$, $p<0.001^*$. For this vowel contrast, all three children's groups also showed poorer performance for the 50% cough stimuli than for whole syllables: 3 year olds, $t(9)=4.03$, $p=0.003$; 5 year olds, $t(14)=6.31$, $p<0.001^*$; and 7 year olds, $t(17)=5.67$,

$p<0.001^*$. Five and 7 year olds also performed more poorly for the reiterated syllable stimuli than for isolated vowels, 5 year olds, $t(14)=3.36$, $p=0.005$; and 7 year olds, $t(17)=4.78$, $p<0.001^*$. In this case, the failure to find a similarly significant effect for 3 year olds is likely due to fewer degrees of freedom: For all three children's groups there was a 4 percentage point difference between conditions.

IV. DISCUSSION

This experiment was conducted primarily to examine whether children rely more on static or dynamic spectral structure in vowel perception, and whether their perceptual strategies for vowels differ from those of adults. Stimuli were designed to differ in which aspect of vowel quality was manipulated within a contrast (height or frontedness) and in the extent of consonant-vowel coarticulation. In some cases, only flat formants heard as isolated vowels were presented to listeners and in other cases only formant transitions heard as syllables, either with or without overlaid coughs, were presented to listeners. Overall results showed that listeners were able to recognize vowels with either static or dynamic spectral information, supporting the conclusion of Diehl *et al.* (1981), but not without possible caveats. In particular, earlier experiments showing a preference in adult listeners for dynamic signal components included more vowels in each contrast (e.g., Jenkins *et al.*, 1983; Strange *et al.*, 1983, 1976), and so performance was generally lower than what was found here. That makes it difficult to compare adults' results from this experiment to those of earlier experiments. With adults' performance consistently close to 100% correct, our ability to observe potential differences across syllable types was constrained. But the focus of this experiment was not on how adults would do, but rather on how children would perform, compared to adults. The finding that listeners of all ages performed as well as they did with only dynamic syllable structure suggests that this structure informs listeners substantially about vowel quality. Certainly this result argues against the notion that children rely more than adults on "…the formant steady states for vowel identification in difficult conditions," as suggested by Sussman (2001, p. 1179). If that were the case we would have expected children to do far worse than adults explicitly with the dynamic stimuli that had syllable centers replaced with a cough. Instead, children were somewhat less accurate than adults, but this age-related discrepancy in performance was not specific to stimuli preserving only dynamic structure. All listeners used the entire spectral pattern for vowel labeling, as long as some minimal amount was available: Listeners had difficulty when only three pitch periods on either side of the cough were heard, which is reasonable given that the very notion of a "dynamic" property suggests that a sufficiently long stretch of signal must be input in order for a perceiver to recognize the change. In fact it was precisely the "formant steady states" in the difficult condition that children had particular difficulty with (i.e., the reiterated pitch periods from the /dæd/ vs /dʌd/ contrast).

The finding that children were more affected than adults by the vowel contrast being heard when stimuli were flat, rather than dynamic, indicates how strongly children rely on

dynamic spectral structure to recognize vowels. With the dynamic signals, children showed similar trends across contrasts to those of adults. With the flat stimuli, children performed much more poorly with the contrast involving tongue frontedness in the /dVd/ context than with the other contrasts. This trend was found even for the natural vowels produced in isolation: Apparently, children were so hindered by hearing only the flat formants that their abilities to recognize vowels in the absence of consonantal contexts were generally disturbed. This result demonstrates how important dynamic signal components are for young children listening to speech.

In general, these findings supported the conclusions of others asserting that vowel identity is specified in the dynamic spectral structure of the entire syllable (e.g., Jenkins *et al.*, 1983; Strange *et al.*, 1983; 1976). At the same time, this study significantly extends that work with adults by demonstrating that children similarly use the dynamic spectral structure of whole syllables to make vowel decisions. Although Murphy *et al.* (1989) had demonstrated this result, the present study uncovered it for a greater number of contrasts, and served to mediate between the contradictory findings of Murphy *et al.* and Sussman (2001). With a contrast involving greater coarticulation between consonant and vowel than those used in either of those studies, and stimuli constructed to be as natural as possible, it became clear that children can and do use information arising from coarticulated segments to make vowel decisions.

This study was unable to provide specific support for Nearey and Assmannn's (1986) claim that vowel-inherent spectral change is critical to vowel recognition, but instead found further support for the claim that dynamic spectral structure arising from articulatory movements between consonant and vowel constrictions is most important. Of course, there has accumulated over the years substantial evidence showing how important dynamic structure is to speech perception, for both consonant and vowel recognition. A few of the many examples regarding consonant perception include a study by Harris (1958) on fricative perception, one by Kewley-Port *et al.* (1983) on syllable-initial stop perception, and one by Nittrouer (2004) on syllable-final stop perception. In all these studies, dynamic structure spanning some portion of the syllable was found to be more important to decisions of consonant identity than either static spectral or temporal structure. If one cobbles together the conclusions of these many studies a theoretical framework emerges based on the idea that dynamic structure in the speech signal plays a fundamental role in phonetic recognition. It is not difficult to fit Nearey and Assmann's finding into that framework: Given a large stimulus set of many spectrally similar vowels, as Nearey and Assmannn used, dynamic structure inherent in the production of isolated vowels would be expected to provide information to help disambiguate between those vowel choices.

Finally, it is worth emphasizing that what is generally termed the "steady-state" syllable region is actually part of the whole dynamic structure of that syllable. When the beginning and ending dynamic components of a syllable are presented to listeners in normal temporal proximity, it is likely the case that listeners "hear through" the silence (or white noise or cough, whatever the case may be). These stimulus portions provide information about both vowel and consonant. But upon hearing only brief, steady-state portions of the vowel target, we would not expect listeners to recognize syllable margins in that same way because there is not enough information available about what consonants were at those margins. With this in mind, the current study serves as a reminder that human speech perception does not unfold as a process of extraction of static spectral slices from the acoustic speech signal, matched against internally stored templates of individual phonemes or features. Rather, speech perception very much depends upon the dynamic structure of the signal, spanning temporal slices generally affiliated with several phonetic segments.

Appendix: Percentages of correct vowel recognition, in each condition for each contrast.

	Age	Isolated vowel	Reiterated vowel	Reiterated syllable	Whole syllable	50% cough	Three pitch periods
/bɪb/ vs /bæb/	3	98.83 (2.86)	98.17 (2.99)	97.83 (2.04)	98.67 (1.73)	94.44 (9.99)	86.22 (13.11)
	5	99.07 (1.94)	98.93 (2.09)	98.21 (2.67)	99.29 (1.49)	96.86 (4.20)	86.14 (11.14)
	7	99.85 (0.55)	99.38 (0.96)	98.85 (2.15)	99.69 (0.75)	98.00 (4.18)	89.38 (13.38)
	Adult	100.00 (0.00)	99.33 (1.30)	98.67 (2.10)	99.83 (0.58)	100.00 (0.00)	99.42 (1.51)
/bæb/ vs /bʌb/	3	95.67 (3.65)	95.42 (4.54)	96.58 (4.38)	99.00 (1.41)	97.90 (2.13)	85.40 (9.24)
	5	95.77 (4.48)	96.00 (4.22)	95.46 (6.10)	99.54 (0.88)	98.54 (2.47)	91.46 (4.20)
	7	99.67 (0.78)	99.33 (1.30)	99.83 (0.58)	100.00 (0.00)	99.67 (0.78)	96.17 (3.95)
	Adult	99.67 (0.78)	99.50 (0.90)	99.33 (1.56)	100.00 (0.00)	99.83 (0.58)	98.83 (1.34)

	Age	Isolated vowel	Reiterated vowel	Reiterated syllable	Whole syllable	50% cough	Three pitch periods
/dɪd/ vs /dæd/	3	95.56 (4.03)	95.00 (5.94)	94.11 (3.95)	96.78 (4.74)	91.78 (6.08)	78.67 (11.70)
	5	99.50 (1.24)	99.33 (1.30)	97.83 (2.59)	99.07 (1.69)	91.57 (10.21)	73.93 (16.00)
	7	100.00 (0.00)	99.83 (0.58)	99.33 (1.56)	99.50 (0.90)	92.83 (10.78)	70.42 (18.11)
	Adult	99.83 (0.58)	99.25 (1.76)	97.92 (4.06)	100.00 (0.00)	99.67 (0.78)	93.75 (5.59)
/dæd/ vs /dʌd/	3	83.44 (19.04)	85.11 (16.31)	79.22 (18.79)	96.40 (3.86)	84.40 (13.62)	52.00 (5.52)
	5	92.80 (11.31)	96.40 (6.29)	88.80 (10.35)	98.67 (2.13)	93.00 (4.14)	56.07 (5.16)
	7	98.72 (2.11)	99.78 (0.65)	94.61 (4.60)	99.22 (1.26)	95.11 (5.05)	56.83 (8.81)
	Adult	99.25 (1.29)	99.42 (1.24)	96.83 (3.69)	99.92 (0.29)	98.08 (2.71)	75.58 (8.36)

ACKNOWLEDGMENTS

[1]Sussman (2001) also included 5- and 6-year-old children with specific language impairments, but results for those listeners are not discussed here.

[2]In spite of frequent attempts to do so, we have never been able to train 3 year olds younger than 3 years, 5 months to perform a labeling task reliably, and so for 3 year olds we use children toward the higher end of their age bracket.

[3]It has been common practice in this laboratory to require data from any one listener to meet specific criteria in order for those data to be included in the final analyses; typically that criterion is that the best exemplars of the stimuli be labeled with 80% accuracy. That sort of criterion is established to ensure that all listeners, particularly children, maintain general attention during testing, and so any observed age-related differences cannot be attributed to a generalized failure on the part of children to stay on task. An explicit criterion was not established in this experiment, other than the pretest criterion, because one goal was to examine labeling accuracy across stimulus types and conditions. However, the fact that all listeners responded with well above 80% accuracy to whole-syllable stimuli, which might be considered the best exemplars in this experiment, is an indication that all listeners maintained general attention to the task during testing.

Cohen J. (**1988**). "Statistical power analysis for the behavioral sciences," Erlbaum (2nd ed.) Hillsdale, New Jersey.

Assmann, P. F., and Katz, W. F. (**2000**). "Time-varying spectral change in the vowels of children and adults," J. Acoust. Soc. Am. **108**, 1856–1866.

Diehl, R. L., McCusker, S. B., and Chapman, L. S. (**1981**). "Perceiving vowels in isolation and in consonantal context," J. Acoust. Soc. Am. **69**, 239–248.

Fant, G. (**1973**). *Speech Sounds and Features* (MIT, Cambridge, MA).

Ferrand, C. T. (**2007**). *Speech Science: An Integrated Approach to Theory and Clinical Practice* (Allyn and Bacon, Boston).

Fox, R. A. (**1989**). "Dynamic information in the identification and discrimination of vowels," Phonetica **46**, 97–116.

Fry, D. B., Abramson, A. S., Eimas, P., and Liberman, A. M. (**1962**). "The identification and discrimination of synthetic vowels," Lang Speech **5**, 171–189.

Gerstman, L. J. (**1968**). "Classification of self-normalized vowels," IEEE Trans. Audio Electroacoust. **AU-16**, 78–80.

Goldman, R., and Fristoe, M. (**2000**). *Goldman Fristoe 2: Test of Articulation* (American Guidance Service, Circle Pines, MN).

Harris, K. S. (**1958**). "Cues for the discrimination of American English fricatives in spoken syllables," Lang Speech **1**, 1–7.

Jastak, S., and Wilkinson, G. S. (**1984**). *The Wide Range Achievement Test-Revised* (Jastak Associates, Wilmington, DE).

Jenkins, J. J., Strange, W., and Edman, T. R. (**1983**). "Identification of vowels in 'vowelless' syllables," Percept. Psychophys. **34**, 441–450.

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.

Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (**1957**). "The discrimination of speech sounds within and across phoneme boundaries," J. Exp. Psychol. **54**, 358–368.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., and Carter, B. (**1974**). "Explicit syllable and phoneme segmentation in the young child," J. Exp. Child Psychol. **18**, 201–212.

Lindblöm, B. E., and Studdert-Kennedy, M. (**1967**). "On the role of formant transitions in vowel recognition," J. Acoust. Soc. Am. **42**, 830–843.

Murphy, W. D., Shea, S. L., and Aslin, R. N. (**1989**). "Identification of vowels in 'vowel-less' syllables by 3-year-olds," Percept. Psychophys. **46**, 375–383.

Nearey, T. M., and Assmann, P. (**1986**). "Modeling the role of inherent spectral change in vowel identification," J. Acoust. Soc. Am. **80**, 1297–1308.

Neely, S. T., and Peters, J. E. (**1992a**). *SPECTO User's Guide* (Boys Town National Research Hospital, Omaha, NE).

Neely, S. T., and Peters, J. E. (**1992b**). *WavEd User's Guide* (Boys Town National Research Hospital, Omaha, NE).

Nittrouer, S. (**2004**). "The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults," J. Acoust. Soc. Am. **115**, 1777–1790.

Routh, D. K., and Fox, B. (**1984**). "'Mm…is a little bit of May:' Phonemes, reading, and spelling," Advances in Learning and Behavioral Disabilities **3**, 95–124.

Strange, W., Jenkins, J. J., and Johnson, T. L. (**1983**). "Dynamic specification of coarticulated vowels," J. Acoust. Soc. Am. **74**, 695–705.

Strange, W., Verbrugge, R. R., Shankweiler, D. P., and Edman, T. R. (**1976**). "Consonant environment specifies vowel identity," J. Acoust. Soc. Am. **60**, 213–224.

Sussman, J. E. (**2001**). "Vowel perception by adults and children with normal language and specific language impairment: Based on steady states or transitions?," J. Acoust. Soc. Am. **109**, 1173–1180.

Syrdal, A. K., and Gopal, H. S. (**1986**). "A perceptual model of vowel recognition based on the auditory representation of American English vowels," J. Acoust. Soc. Am. **79**, 1086–1100.

Walley, A. C., Smith, L. B., and Jusczyk, P. W. (**1986**). "The role of phonemes and syllables in the perceived similarity of speech sounds for children," Mem. Cognit. **14**, 220–229.

Warren, R. M. (**1970**). "Perceptual restoration of missing speech sounds," Science **167**, 392–393.

Warren, R. M., and Obusek, C. J. (**1971**). "Speech perception and phonemic restorations," Percept. Psychophys. **9**, 358–362.

Wightman, F., and Kistler, D. (**2005**). "Informational masking of speech in children: Effects of ipsilateral and contralateral distracters," J. Acoust. Soc. Am. **118**, 3164–3176.

A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition

Laura C. Dilley[a) and Mark A. Pitt

Department of Psychology, Ohio State University, 1835 Neil Ave., Columbus, OH 43220

Regressive place assimilation is a form of pronunciation variation in which a word-final alveolar sound takes the place of articulation of a following labial or velar sound, as when *green boat* is pronounced *greem boat*. How listeners recover the intended word (e.g., *green*, given *greem*) has been a major focus of spoken word recognition theories. However, the extent to which this variation occurs in casual, unscripted speech has previously not been reported. Two studies of pronunciation variation were conducted using a spontaneous speech corpus. First, phonetic labeling data were used to identify contexts in which assimilation could occur, namely, when a word-final alveolar stop (/t/, /d/, or /n/) was followed by a velar or labial consonant. Assimilation was indicated relatively infrequently, while deletion, glottalization, or canonical pronunciations were more often indicated. Moreover, lexical frequency was shown to affect pronunciation; high frequency lexical items showed more types of variation. Second, acoustic analyses showed that neither place of articulation cues (indicated by second formant variation) nor relative amplitude was sufficient to distinguish assimilated from deleted and canonical variants; only when closure duration was additionally taken into account were these three variant types distinguishable. Implications for theories of word recognition are discussed. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2772226]

I. INTRODUCTION

There can be a great deal of variability in how words are pronounced (e.g., Dalby, 1986; Bell *et al.*, 2003; Shockey, 2003; Johnson, 2004), yet listeners rarely experience difficulty in understanding what is said. One type of pronunciation variation that occurs in connected speech is regressive place assimilation, in which the final alveolar segment of a word is produced at the same place of articulation as a following segment. For instance, the /n/ at the end of *green* may take the labial place of the following /b/ in the phrase *green boats*, so that *green* appears to be pronounced as *greem*.

How do listeners recognize the intended word (i.e., *green* given *greem*) when assimilation occurs? This question has received considerable attention in the experimental literature, with multiple theoretical accounts being put forth (e.g., Lahiri and Marslen-Wilson, 1991; Gaskell and Marslen-Wilson, 1998; Gow, 2003). Although this work has been informed by a large body of research on the articulatory and acoustic characteristics of assimilation (e.g., Wright and Kerswill, 1989; Holst and Nolan, 1995; Zsiga, 1995; Ellis and Hardcastle, 2002), there remain unanswered questions about the scope and nature of the problem that assimilation poses for recognition. For example, at a phonological level, little is known about the frequency of assimilation relative to other types of pronunciation variation that might also occur in contexts in which assimilation is possible (i.e., in assimilable environments). Furthermore, most articulatory and acoustic studies have examined assimilation in read speech,

raising questions about how representative these data are of the unscripted, conversational speech to which listeners are most frequently exposed. Thus, understanding the extent to which assimilation and other connected speech processes occur in spontaneous speech is necessary to ensure theories of spoken word recognition can adequately and accurately account for how pronunciation variants are recognized.

The use of speech corpora to inform theorizing and experimentation in speech perception and production has been increasing in recent years. One early study by Dalby (1986) investigated effects of speaking rate on the likelihood of deleting schwa vowels in American English. The results showed schwas deleted more often at fast speech rates and in word-medial position, but no differences were found as a function of lexical stress environment. Subsequently, Patterson *et al.* (2003) examined frequency of schwa deletion in conversational American English. They found that lexical stress environment was the most important factor in predicting deletion of schwa vowels, in contrast to Dalby (1986). [See Crystal and House (1988) for a comparison of stress and speech rate effects on segment realizations in speech corpora.]

Moreover, Patterson and Connine (2001) investigated frequency of allophonic variants of word-medial /t/ in conversational American English. They showed that lexical items which were high frequency and/or less complex morphologically were more likely to show /t/ realized as the flapped variant [ɾ].[1] In Dutch, Mitterer and Ernestus (2006) used read and spontaneous speech corpora to investigate optional word-final /t/-lenition. They found that /t/-lenition occurred more often in spontaneous than read speech, but could identify no context which consistently induced /t/-lenition.

[a)]Electronic mail: dilley@bgsu.edu. The author is now with the Department of Communication Disorders and Department of Psychology, Bowling Green State University, 247 Health Center, Bowling Green, OH 43403.

Additional studies on Dutch have shown that word frequency influences the likelihood of voicing assimilation (Ernestus *et al.*, 2006) and of short durations for affixes (Pluymaekers *et al.*, 2005). Moreover, the studies by Ernestus *et al.* (2006) as well as Snoeren *et al.* (2006) for French each suggest that voicing assimilation is graded, rather than categorical; both studies used read speech. Among other things, these results highlight the fact that speech style (e.g., read versus spontaneous) affects rates of variant modifications. The current research builds on this work by analyzing the types of variation occurring in environments where regressive place assimilation is phonologically possible in spontaneous, unscripted American English.

Three key findings about regressive assimilation have emerged from the empirical literature, almost all of which has used read speech. First, an alveolar stop which assimilates to a following labial or velar consonant often shows acoustic or articulatory evidence of both alveolar place of articulation, as well as labial or velar places of articulation, respectively (e.g., Kohler, 1990; Ohala, 1990; Barry, 1992; Ellis and Hardcastle, 2002; Gow, 2001, 2002, 2003). For example, electropalatography (EPG) and electromagnetic articulography data have demonstrated evidence of partial alveolar assimilations in alveolar-velar (e.g., /d#g/) sequences, indicating a tongue blade/body gesture along with velar contact of the tongue dorsum (Wright and Kerswill, 1989; Barry, 1992; Nolan, 1992; Zsiga, 1995). Similarly, acoustic studies have demonstrated that for assimilated alveolar stops followed by labials, the mean formant frequencies are intermediate between those of unassimilated alveolar and labial sounds (Gow, 2001, 2002, 2003).

A second finding is that assimilation often leads to gradient cues to place of articulation of the word-final segment (Gow, 2001, 2002, 2003; Holst and Nolan, 1995; Nolan *et al.*, 1996). For example, in an EPG study of alveolar-velar sequences (e.g., *road collapsed*) Wright and Kerswill (1989) found that speakers produce varying alveolar and velar contact, ranging from full alveolar with no velar contact, to a mixture of alveolar and velar contact, to full velar with no alveolar contact. This gradience is at least partially due to the fact that the degree of assimilation produced often varies both within and across talkers (e.g., Nolan *et al.*, 1996; Ellis and Hardcastle, 2002). That this gradience is relevant for processing has been demonstrated through perceptual studies of assimilation cues (Nolan, 1992; Gow and McMurray, in press). Such findings are reminiscent of work showing variability within phonetic categories and listener sensitivity to this variation (e.g., Miller, 2001), as well as variability in phonetic realizations as a function of lexical items, e.g., the tendency in American English to produce /t/ as [ɾ] in *pretty* (Patterson and Connine, 2001; Connine, 2004).

A third finding is that in cases where alveolar assimilation is extreme, assimilated forms and underlying nonalveolar forms can be indistinguishable (Holst and Nolan, 1995; Nolan *et al.*, 1996). For example, Holst and Nolan (1995) found on the basis of acoustic measurements that the assimilation of /s/ to [ʃ] was rarely differentiable from the underlying form (/ʃ/ to [ʃ]). Other studies have replicated this result using articulatory data (Nolan *et al.*, 1996; Ellis and Hardcastle, 2002).

These three findings collectively suggest that except in cases of the most extreme alveolar assimilation, assimilated segments are likely differentiable acoustically and/or articulatorily from canonical forms. A few perceptual studies have shown listeners can distinguish between alveolars realized as assimilated versus unassimilated as well (e.g., Nolan, 1992; Gow and McMurray, in press). However, the usefulness of remnant cues to underlying alveolar place for assimilated consonants in perception depends in part on these being reliably present. In this regard, it is not clear whether in spontaneous, unscripted speech alveolars in assimilable environments will retain remnants of their underlying place of articulation. Previous work suggests that place assimilation is more common during casual speaking styles, as well as at faster speech rates (e.g., Barry, 1992); however, these results were obtained using scripted speech. Spontaneous speech is known to show different acoustic-phonetic attributes from read speech, including greater gestural overlap, a higher degree of segmental deletion, and different strength of consonantal gestures (e.g., Browman and Goldstein, 1990; Johnson, 2004; Shockey, 2003).

Our goal in the present investigation was to build on prior research by examining a corpus of spontaneous speech to determine for assimilable environments the extent to which assimilation versus other forms of pronunciation variation occurred. This broad perspective enabled us to define and compare assimilation relative to other forms of word-final variation. These data should in turn inform theorizing about recognizing assimilated word forms by clarifying the challenges that the perceptual system faces.

There were two parts to the study. First, we examined the relative frequencies of distinct phonetic variants occurring in assimilable environments as given by phonetic labels in the speech corpus, in order to determine the consistency with which assimilation occurs. Second, we evaluated the extent to which assimilated alveolars are acoustically differentiable from unassimilated alveolars, labials, and velars. To ensure the generality of the results, we examined variation in multiple word-final alveolar segments (/t/, /d/, and /n/) in the context of a following velar (/g/ or /k/) or labial (/p/, /b/, or /m/) segment. In addition, this was done when the context preceding the alveolar was labeled as a high front vowel ([ɪ]) and a nonhigh front vowel ([æ] or [ɛ]).

II. PHONETIC LABELING ANALYSIS

A. Method

The relative frequency of regressive place assimilation was investigated using phonetic labels from the Buckeye Corpus of Conversational Speech (Pitt *et al.*, 2006); this corpus is comprised of conversations from talkers in the Columbus, OH area. Phonetic transcriptions were made by a group of trained phonetic labelers, who were paid for corpus preparation. Labelers used spectrogram and waveform displays in Xwaves software (Entropies Corp.), as well as auditory cues, to label the phonetic segments present in the speech accord-
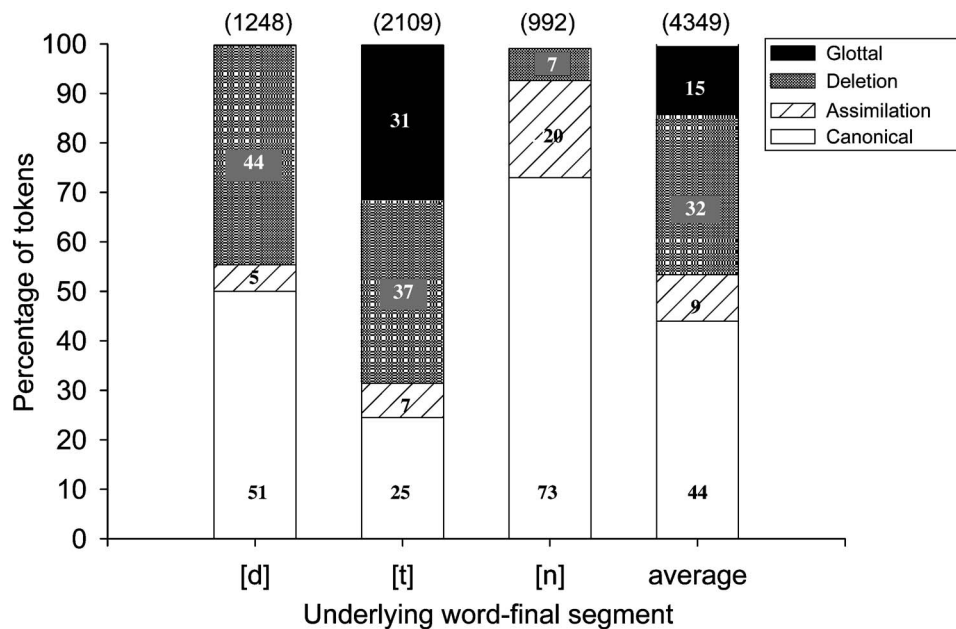
FIG. 1. Percentage of tokens labeled as each of the four variants for the three word-final alveolars in an assimilable context. The number of tokens of each alveolar is listed in parentheses.

ing to the labeling conventions described in Kiesling *et al.* (2006). Of relevance to the present study were conventions for labeling segments as assimilated, deleted, glottalized, and canonical. Labelers indicated a segment to be *assimilated* when (1) the perceptual evidence was consistent with a word-final alveolar stop adopting a labial or velar place of articulation, and (2) there was spectral evidence of a local change in F2 in a sonorant segment just preceding the stop closure (i.e., F2 fell or rose in the case of an alveolar stop adopting a labial or a velar place or articulation, respectively). Moreover, labelers indicated a segment to be *glottalized* when the segment had perceptually creaky voicing accompanied by irregularity in pitch period timing in the waveform. (See Dilley *et al.*, 1996; Redi and Shattuck-Hufnagel, 2001.) In addition, labelers indicated a segment to be *deleted* when it could not be heard when a short context was played and there was no clear visual evidence in the spectrogram that the segment was present. Finally, labelers indicated a segment to be *canonical* when it was perceived as present and unassimilated, lacked creaky voicing, and/or when they were uncertain about the variant type.[2]

Tests of transcription consistency and agreement among labelers were performed periodically during creation of the corpus, and have indicated a high degree of reliability in the use of the phonetic labels. In particular, in a published study of intertranscriber reliability (Pitt *et al.*, 2005) using four labelers annotating 4 min of speech, overall agreement for all phonetic labels was 80.3%, with agreement of 92.9% for stops. A more recent unpublished test of intertranscriber reliability using eight labelers and 1 min of speech allowed us to specifically investigate agreement among canonical, deleted, and glottal variants. Agreement for these variants was 85.2%, indicating high reliability in line with previous findings of good interrater agreement (e.g., Irwin, 1970; Eisen, 1991).[3]

The speech of 19 talkers (9 male, 10 female; approximately 138 000 words) was used to identify lexical sequences constituting assimilable environments, i.e., two-word sequences in which the place of articulation of the word-final phoneme could assimilate to that of a following word-initial phoneme. Analyses were limited to word-final alveolars (/t/, /d/, or /n/) that were followed by word-initial labials (/b/, /p/, or /m/) or velars (/g/ or /k/), since these environments are subject to processes of place assimilation in English (Shockey, 2003). These phonological environments were identified using citation pronunciations obtained from a phonetic dictionary (CMU pronouncing dictionary 0.6), using the orthographic transcriptions of the conversations. Note that two-word sequences which were labeled by phonetic analysts as showing a fluent or nonfluent pause at the word boundary were eliminated from analysis. [See Kiesling *et al.* (2006) and footnote 2 for more details.]

To determine what phonetic changes occurred in assimilable environments, the "underlying" or citation pronunciations were compared with the "surface" or actual pronunciations of the words in the corpus. The four surface forms that underlying word-final alveolar sounds could take based on transcription conventions were assimilated, deleted, glottalized, or canonical. Finally, note that word-final alveolars could not be realized as a flap in the contexts under investigation, because such variations are limited to positions between vocalic or sonorant segments (De Jong, 1998).

B. Results and discussion

There were a total of 4349 assimilable contexts. These are shown in Fig. 1, which gives the rate of each of four possible variant realizations (assimilation, deletion, glottal, canonical) as a function of underlying segment type (/t/, /d/, or /n/). The number of tokens included for each of the three

TABLE I. Percentage of tokens in assimilable environments as a function of form class, number of syllables, and assigned label. All percentages have been rounded.

	Function word		Content word	
	Monosyllabic	Polysyllabic	Monosyllabic	Polysyllabic
Canonical	15	3	18	6
Assimilation	5	0	4	1
Deletion	15	4	9	4
Glottal	9	1	4	0
SUM	44	8	35	11

identity (/t, d, n/) because all showed a consistent pattern at this level of analysis. The data are shown in Table I.

Comparison of the totals in the last row shows that function and content words occur equally often in assimilable environments. Given the small number of function words in English relative to content words, these data partially forecast what will become clear shortly, that a small number of function words make up the majority of items in assimilable environments. Table I also reveals usage statistics in the language: Monosyllables make up a higher percentage of assimilable tokens for both word form types, consistent with the fact that monosyllables make up 80% of the tokens in the Buckeye Corpus (Pitt *et al.*, 2005).

More interesting in Table I, however, is how the percentages of tokens differ across pronunciation variants. For monosyllabic function words, for instance, deletions are as frequent as canonical pronunciations. In contrast, monosyllabic content words show a canonical bias of approximately 2:1 relative to deletions. This asymmetry in frequency of canonical versus deleted realizations is present for polysyllabic words as well, only to a much smaller degree. Thus, function words appear more likely to deviate from their canonical pronunciations than content words. [See Raymond *et al.* (2006) for a similar finding.] Moreover, rates of assimilation do not differ appreciably for monosyllabic function and content words; both are relatively rare. Finally, glottal realizations were more common for function words than for content words.

The next analysis examined the frequency of different variant realizations for individual words represented in Table I. Given a particular word, was that word more likely to be realized in mainly one way (e.g., canonical only) or in more than one way (e.g., both canonical and deleted)? The data are shown in Table II, with the upper half reflecting word types and the lower half number of tokens. The values in each cell are percentages of the total number of types or tokens. The column labels designate the type(s) of variation exhibited by each word final segment (D=deletion, A=assimilation, C=canonical, G=glottal), with multiple letters indicating that the word-final segment showed each of those realizations at least once. Because the segments /d/ and /n/ were

segments is in parentheses at the top of the graph. The canonical variant constituted the most frequent type of surface realization, occurring 44% of the time overall. Noncanonical surface realizations were quite frequent as well, constituting 27%–75% of instances across the three segments. Assimilation was indicated infrequently, occurring only 9% of the time across the three segmental environments; it was least common for /t/ and /d/ (5% and 7%, respectively) and most common for /n/ (20%). Deletions were also quite common, especially for the oral stops; they constituted 45% and 37% of the /d/ and /t/ realizations, respectively. Finally, glottal variants were found almost exclusively for /t/, for which they occurred almost as often as deletions (31%).

The high rate of deletion for /t/ and /d/ was somewhat unexpected. It has been reported that /t/ and /d/ readily delete in the context of a preceding /n/ both word-medially (Raymond *et al.*, 2006) and word-finally (Guy, 1980; Neu, 1980). Consistent with this earlier work, a high percentage (53%, $N=1408$) of deleted /t/ and /d/ tokens were found to have been preceded either by /n/ or syllabic /n/. No other systematicities were identified in the remaining cases of deletion.

The data in Fig. 1 were analyzed further to provide a more complete picture of phonological variation in assimilable environments. We began by calculating the frequency of the four realizations as a function of word class (function or content word) and length. Monosyllables, which constituted 79% of the tokens, were compared with polysyllabic words. The data were combined over underlying segment

TABLE II. Percentage of word types (upper half) and word tokens (lower half) labeled as having been realized as each of four variant categories (D=deletion, A=assimilation, C=canonical, G=glottal), broken down by final segment (/t/, /d/, or /n/). Categories with two or more labels (e.g., DA) indicate the word-final segment was indicated as having been realized as more than one variant type. Note that possible categories for /t/ are listed separately because glottals rarely occurred for /d/ or /n/. See the text for further details.

	Labeling categories							
Types	D	A	C	DA	DC	AC	DAC	
/d/	8	5	60	2	8	9	8	
/n/	4	10	58	3	3	14	9	
	DG	AG	CG	DAG	DCG	ACG	DACG	G
/t/	32	5	29	1	22	2	9	1
Tokens	D	A	C	DA	DC	AC	DAC	
/d/	6	1	15	1	35	12	32	
/n/	1	2	14	1	2	15	65	
	DG	AG	CG	DAG	DCG	ACG	DACG	G
/t/	7	1	5	1	22	1	65	0

almost never labeled as glottal variants (occurring just twice for /d/), only seven categories are given for these phones, as indicated by seven columns for single or multiple realization types. In contrast, /t/ was additionally labeled as a glottal variant very frequently, so up to 15 combinations of single or multiple realizations were possible. However, glottal variants were almost always a possible realization of /t/-final words in addition to at least one other variant type, so that many possible cells involving single realization types for /t/ were empty. Thus, only 8 of 15 possible combinations of realizations are shown in Table II (since the other possible combinations had *N*=0). In many ways, the results across categories were similar for /t/-final words as for /d/ and /n/-final words, except for the additional glottal variation.

Consider the data in the first three columns in which a variant was realized in only one way (two for /t/). For deletions, assimilations, and glottals the percentage of types and tokens is small (less than 14%), indicating few words are spoken only in a reduced form. The one exception is /t/, for which 40% of the /t/-final words were never spoken canonically. For canonical realizations, there is a marked difference in frequency between types and tokens. Type frequency is enormous, greater than 50% in the cases of /d/ and /n/, indicating that a large number of words were spoken only in their canonical form. The corresponding token frequencies are four times smaller, indicating that these words were spoken infrequently in the corpus. Note also that this pattern was greater overall for /d/ and /n/ than for /t/.

The final segment was realized in two ways (three for /t/) for items represented in the middle section of Table II, and the same bias for the canonical pronunciation is present here. Word-final segments that were realized as DA (deleted or assimilated) occurred less often than the combinations of DC and AC, both for types and for tokens. In general, percentages were low across the DA, DC, and AC columns, and there is not much change between type and token values, except for DC. Here, the token frequency for /d/ is fourfold that of type frequency, just the reverse of what was found when only one realization occurred (first three columns). In this case, a small set of words is responsible for a disproportionately large amount of variation. This same pattern of a few types being responsible for most variation is found in the DAC (DACG) column, where words were realized in all three (or four) ways. The result is particularly dramatic for the /n/ and /t/ words. Although they constituted less than 10% of the word types in the corpus, they accounted for almost two-thirds of the tokens that underwent variation.

These data indicate that most of the word-final variation in assimilable environments occurs in a small subset of words whose final segment can assimilate, delete, and in the case of /t/, glottalize. What are these words? Shown in Fig. 2 are the 21 most frequent words in our sample broken down by type of variation. They are grouped by the identity of the final segment. Together, they constitute 67% of the sample (2956 tokens). Most of the 21 are monosyllabic (19) and function words (14). Most of the /d/-final and /n/-final words show evidence of assimilation and deletion, though more of the former, except for *and*, whose final /d/ deletes most often (Neu, 1980). Many /t/-final words are realized in all four
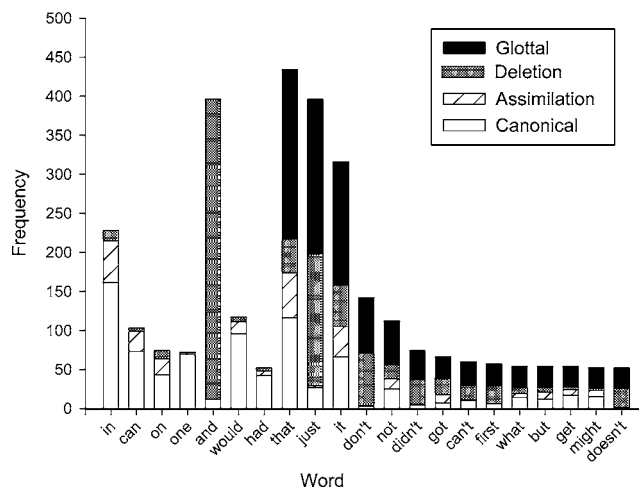
FIG. 2. Variant rates for the 21 most frequent words in the phonetic labeling analysis.

ways. The preponderance of glottal forms for /t/-final words is particularly striking. These data provide a highly representative depiction of the variation found in assimilable environments. Furthermore, it is representative across talkers, too, as most speakers (mean=16.9 out of 19) contributed tokens to the counts of each word.

In sum, the phonetic analysis demonstrates that assimilated segments were heard relatively infrequently in environments where they might be expected to be found. By contrast, other phonetic realizations were more common, with canonical pronunciations and deletions predominating. In addition, variant realization varied by final segment type (/t/, /d/, or /n/). Finally, there was considerable variation according to word types, with a small number of very frequent items exhibiting all forms of variation and many other lower frequency items being realized only canonically.

These results suggest that recognizing a word in an assimilable environment is not just a matter of distinguishing between a surface (assimilated) form and an underlying (canonical) form. The large number of deletions, as well as glottal variants, indicate that the problem of recognition, even in this highly constrained context, is more complex than has often been assumed, and suggests that a high degree of flexibility in variant recognition is necessary for word recovery to succeed. In Sec. III we take a closer look at this variation by analyzing some acoustic characteristics in the vicinity of key word-final segments.

III. ACOUSTIC ANALYSIS

A. Method

Acoustic analyses were undertaken in order to probe further the nature of variation in word-final alveolars in assimilable environments. The second formant (F2) of a vowel is strongly affected by the place of articulation of an upcoming or preceding consonant (Stevens, 1998). Therefore, F2 was measured in the vicinity of the underlying word-final alveolars to estimate the degree to which cues to the place of articulation of the following word-initial consonant were indicated. To ensure comparable phonological contexts, analy-
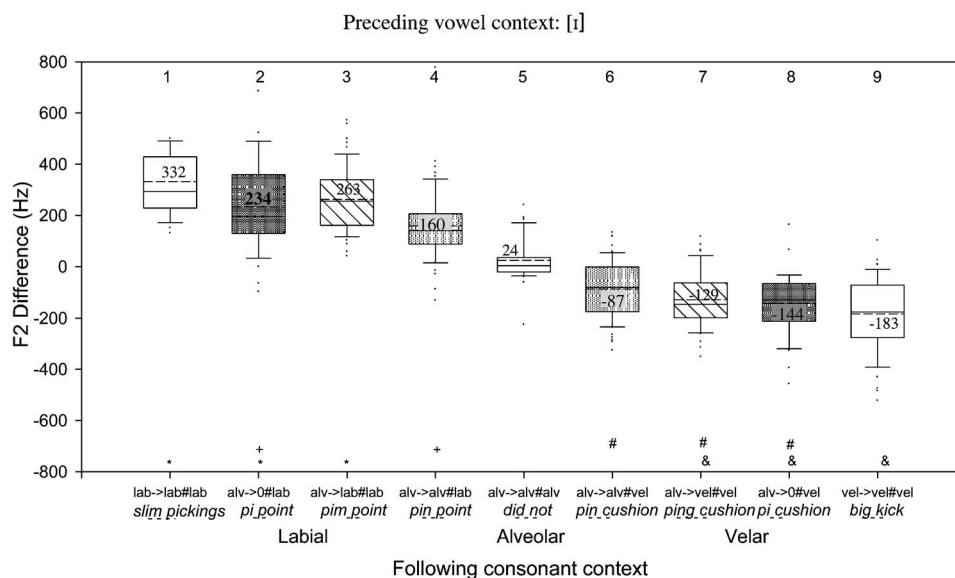
FIG. 3. Box plots of the F2 difference in an [ɪ] vowel show nine labeling conditions in which segments were followed by labial (bars 1–4), alveolar (bar 5), and velar (bars 6–9) contexts. Labels on the *x* axis describe the conditions, as well as how the word final tokens in those conditions were labeled. The underlying segment is listed first, followed after the arrow by its surface (i.e., labeled) realization. The following context is listed after the word boundary symbol (#). lab, alv, and vel refer to labial, alveolar, and velar places of articulation, respectively. 0 denotes the segment was deleted. Conditions with the white bars contain data from control contexts in which the place of articulation across the word boundary is the same. In bars 2–4 and 6–8, an underlying alveolar segment is in the context of a following labial or velar segment, respectively. In these conditions, the light-grey bars represent data from tokens labeled as canonical or assimilated, and the dark grey bars, deletions. The solid line in the middle box is the median of the distribution. The dashed line is the mean, with its value listed. Symbols at the bottom of the graph denote which conditions are statistically indistinguishable from each other.

ses were limited to tokens in which the underlying word-final alveolar was preceded by a vowel. Moreover, because the tongue height and advancement of vowels also affect F2 (Stevens, 1998), the identity of vowels in contexts preceding the underlying word-final alveolars was also controlled. Tokens in two kinds of preceding vocalic environments were selected for analysis: (1) the front high vowel [ɪ] as in *sit*, (2) the front nonhigh vowels [æ] and [ɛ] as in *sat* and *set*. These contexts were selected because they were most numerous in the corpus. Tokens with these vowels were identified using the phonetic labels in the corpus.

For each of these two vowel contexts, F2 difference values were determined for underlying alveolar consonants in a total of six assimilable conditions, arising from three possible segmental realizations for these alveolars (canonical, assimilated, or deleted), given either of two kinds of following word-initial consonants (labial or velar).[4] To assess the degree of F2 variation in underlying alveolar tokens, F2 difference measurements were obtained from vowels in control contexts in three nonassimilable environments which had homorganic places of articulation across the word boundary: Word-final labials preceding labials (e.g., *him pay*), word-final alveolars preceding alveolars (e.g., *did not*), and word-final velars preceding velars (e.g., *anything comes*). This gave nine token conditions for each vowel context, for a total of 737 tokens across both vowel contexts. For the [ɪ] context, the average number of tokens in each condition was 49.7 (range 22–71). For [æ] and [ɛ] contexts, the average number was 32.2 (range 9–50).

To calculate the F2 difference (measured in hertz), F2 was measured at two points: (1) the vowel midpoint and (2) at the final pitch period of the vowel, just before consonantal closure (cf. Pitt and Johnson, 2003). A combination of automatic and hand measurements were used to estimate the F2 difference. Automatic measurements were obtained by extracting formant frequencies using an Xwaves script. Hand measurements were made from spectra generated from a 25 ms Hanning window centered on the zero crossings of the pitch periods closest to the vowel midpoint and vowel end point, and/or by measuring F2 values from wide-band spectrograms.

B. Results

Figure 3 shows box plots of the differences in F2 values for a preceding [ɪ] vowel for the six assimilable and three control conditions. The bars are ordered with respect to place of articulation of the following word-initial consonant, with the data from labial contexts in bars 1–4 and velar contexts in bars 6–9; bar 5 shows the data from the alveolar context. The dark grey bars (2 and 8) represent instances of deletion. The bars with hash marks represent cases in which underlying alveolars were classified as assimilated or canonical. The white bars (1, 5, 9) are the control conditions in which place of articulation was the same underlyingly across the word boundary (e.g., *him pay*). Recall that these were included as referents against which to compare the direction and extent of F2 deviation in the other conditions. The horizontal bar through the middle of each box is the median of the distribution. The dashed line is the mean, with it value given.

First consider the data on the left-hand side of the graph, where segment realization could be influenced by a following labial. For alveolar segments labeled as assimilated (bar 3), the extent of the F2 transitions approaches that of true
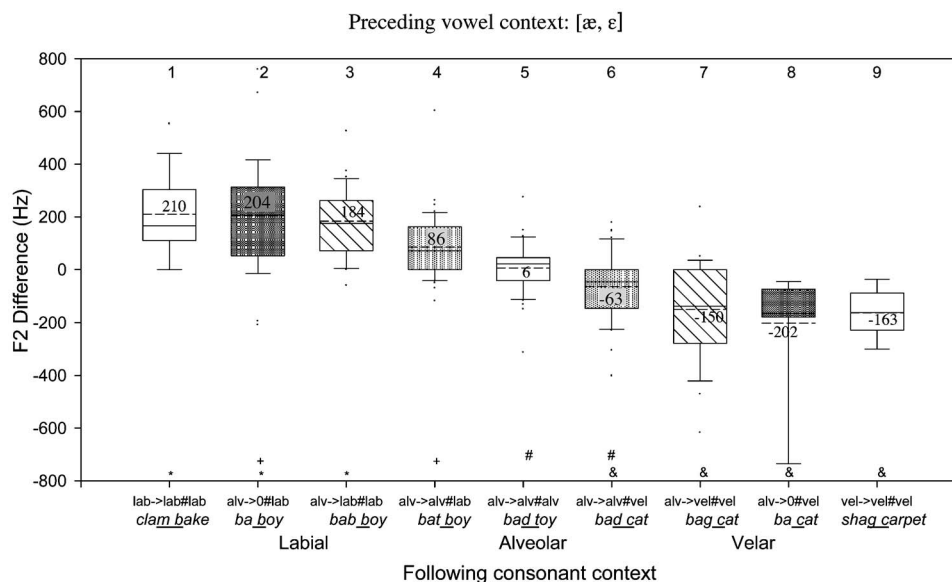
Preceding vowel context: [æ, ɛ]

FIG. 4. Box plots of the F2 difference in [æ] and [ɛ] vowels show nine labeling conditions in which segments were followed by labial (bars 1–4) alveolar (bar 5), and velar (bars 6–9) contexts. Labels on the *x* axis describe the conditions, as well as how the word final tokens in those conditions are labeled. The underlying segment is listed first, followed after the arrow by its surface (i.e., labeled) realization. The following context is listed after the word boundary symbol (#). lab, alv, and vel refer to labial, alveolar, and velar places of articulation, respectively. 0 denotes the segment was deleted. Conditions with the white bars contain data from control contexts in which the place of articulation across the word boundary is the same. In bars 2–4 and 6–8, an underlying alveolar segment is in the context of a following labial or velar segment, respectively. In these conditions, the light-grey bars represent data from tokens labeled as canonical or assimilated, and the dark grey bars, deletions.

labials (bar 1), but falls short on average by 79 Hz. Nevertheless, relative to the alveolar context (bar 5, 239 Hz difference), segments labeled as assimilated are much more similar to true labials. For underlying alveolar segments labeled as deleted (bar 2), not only is the F2 transition affected by the labial place of articulation of the following segment, it is affected to an extent almost as great as segments labeled as assimilated, falling short of this category by a mere 29 Hz. In terms of F2 transition information, assimilated and deleted categories are virtually indistinguishable, as indicated by the fact that their middle quartiles are highly similar. In addition, each distribution overlaps extensively with the labial distribution (bar 1).

Next, for alveolar segments labeled as canonical in the context of a following labial (bar 4), the amount of F2 change is not as great as in assimilated cases (bar 3), as expected. However, it is surprisingly large, with the bulk of the distribution clearly separate from that of the control alveolar context (bar 5). The frequency drop in F2 of these alveolars suggests some degree of assimilation has taken place, but the drop was not enough for labelers to classify the segment as assimilated. These data indicate that a substantial amount of F2 deviation is tolerated before a segment is classified as labial (or missing), despite the fact that this cue is almost certainly being used by labelers to distinguish alveolars from labials (comparing bars 1 and 5).

Similar results are found for alveolars in a following velar context (bars 6–9), but in the opposite direction. For alveolars labeled as having assimilated to a following velar (bar 7), F2 transitions frequently extend as far as true velars (bar 9), falling short on average by 54 Hz. For alveolars labeled as deleted (bar 8), F2 is affected by the place of the following segment to an even greater degree than segments

labeled as assimilated, differing from true velars by 39 Hz. Just as in labial contexts, assimilated and deleted categories in velar contexts appear indistinguishable on the basis of F2 information. In addition, relative to the alveolar context (bar 5), deleted and assimilated segments are much more similar to velars. Finally, for alveolar segments labeled as canonical (bar 6), the extent of F2 change is again large, with the bulk of the distribution offset from that of control alveolar contexts (bar 5), but not entirely overlapping with that of a segment labeled as assimilated (bar 7). The data from velar contexts again suggest that large F2 transitions must be present before an assimilated segment (or no segment) is perceived.

Statistical analyses were carried out to determine which distributions were reliably different from one another. A one-way analysis of variance (ANOVA) showed a significant main effect of condition, $F(8,438)=111.4$, $p<0.01$. The symbols at the bottom of the graph specify which conditions differed reliably in Tukey's honestly significantly different (HSD) post hoc tests. Conditions that share the same symbol at the bottom of the graph are statistically indistinguishable (e.g., 1–3; 2 and 4). The results are virtually symmetrical across the labial and velar contexts. The deleted and assimilated distributions are not only statistically equivalent to each other, but also to the corresponding control (labial and velar) distributions. The main difference across contexts is that the alveolars (bars 4 and 6) are more distinguishable when the following segment is labial than velar.

F2 transition analyses in [æ] and [ɛ] vowel contexts are displayed in Fig. 4; these contexts replicate what was found for the [ɪ] context. For alveolars labeled as having assimilated to a following labial or velar (bars 3 and 7, respectively), the extent of the F2 transitions are comparable to

those of true labials and velars (bars 1 and 9), with the distributions overlapping almost completely. Similarly, for alveolars labeled as having been deleted in the context of a following labial or velar (bars 2 and 8), the distributions overlap fully in the labial context and greatly in the velar context. (Note that the large variability for bar 8 is likely due to the small number of cases of alveolars labeled as having been deleted in the context of velars in our corpus, $n=11$.) Relative to the alveolar context (bar 5), alveolars labeled as assimilated or deleted (bars 2, 3, 7, and 8) are much more similar to labials and velars (bars 1 and 9). These results provide further evidence that assimilated and deleted categories are indistinguishable on the basis of F2 information. Finally, alveolar segments labeled as canonical in labial and velar contexts (bars 4 and 6, respectively) show rather large F2 transitions, such that the bulk of the distributions are separate from that of control alveolar contexts (bar 5). However, these distributions do not entirely overlap with those of segments labeled as assimilated (bars 3 and 7, respectively). This again suggests that the extent of F2 transition can be substantial before a different category is indicated.

A one-way ANOVA across the nine conditions was reliable, $F(8,280)=30.75$, $p<0.01$. Post hoc Tukey's comparisons for labial contexts yielded an almost identical pattern of reliable differences as in Fig. 3. For velar contexts, the differences between distributions were not statistically reliable, probably because of the somewhat greater dispersion of values.

A final analysis of the F2 difference measurements involved combining the data across vowel contexts into a single ANOVA with two variables, preceding vowel context ([ɪ] vs [æ] and [ɛ]) and word-boundary condition (1–9). This analysis was performed mainly to increase the stability of the results, especially in the few cells in which observations were low. The main effect of the word-boundary condition was reliable, $F(8,718)=110.85$, $p<0.001$, and post hoc Tukey's tests showed even more clearly that deletions are indistinguishable from assimilations. Both assimilated and deleted distributions approximate quite closely the control distributions, be they labial or velar; conditions 1–3 were not reliably different, nor were 7–9. Conditions 4–6, in contrast, differed reliably from each other and all other conditions. The only exception to this was condition 6, which was not reliably different from 7. The ANOVA also yielded a significant main effect of vowel context, $F(1,718)=11.857$, $p<0.001$, with F2 differences being smaller overall for [æ] and [ɛ] than [ɪ]. The interaction of the two variables was also reliable, $F(8,718)=2.30$, $p<0.02$, with the F2 difference being smaller in [æ] and [ɛ] contexts when the following environment was labial, but much more comparable to [ɪ] when the environment was velar.

The results of the F2 analyses are quite consistent, replicating across two different preceding vowel contexts. Moreover, by detailing variation in informal, unscripted speech, the present results provide a snapshot of the kinds of variation that listeners typically encounter in conversation, and thus what models must account for in explaining how spoken words are recognized. These findings agree with studies using casually produced read speech, which have shown that assimilated alveolars often exhibit values in place of articulation metrics which are intermediate between alveolar and nonalveolar place (e.g., Gow, 2001, 2002, 2003). The present results for spontaneous speech therefore extend and validate previous studies using read speech.

In addition, these results increase our understanding of pronunciation variation in assimilable environments by demonstrating comparable degrees of acoustic modification of place information for perceptually distinctive phonetic realizations, i.e., assimilations versus deletions. In particular, measurements of F2 difference were indistinguishable for deleted and assimilated categories; these two in turn were statistically indistinguishable with respect to three of four control conditions (i.e., true labial or velar contexts). Only in the labial context in Fig. 3 is the control distribution noticeably shifted upward away from the assimilated and deleted distributions.

The data in the four deleted conditions are particularly surprising because labelers listen closely to the speech for evidence of segments, and they adopt a conservative criterion when classifying a segment as deleted (Kiesling *et al.*, 2006). Although the extent of F2 transitions were comparable for both deletions and assimilations, they were apparently not perceptible in the former case. Why? Analyses reported in the next sections were carried out to identify acoustic evidence that labelers might have used in distinguishing deletions from assimilations.

C. Amplitude differences in formant transitions

One possible reason why labelers reported a word-final segment to be deleted rather than assimilated could have been that the amplitude of the F2 transitions decreased more dramatically in the former case than the latter, making F2 cues less perceptible. We evaluated this idea by measuring amplitude in the vicinity of the underlying segment for a subset of tokens labeled as deleted and assimilated in assimilable environments; measurements from alveolar segments labeled as canonical in assimilable environments were also included for comparison. To ensure comparable samples across these categories, tokens were matched according to several dimensions, including the identity, gender, and/or age of the talker, the type of preceding vowel context, and the place of articulation of the following context (as labial or velar).[5] For [ɪ] contexts, 78 assimilated, 78 deleted, and 78 canonical tokens were examined, while for [æ] and [ɛ] contexts, 45 assimilated, 45 deleted, and 45 canonical tokens were examined.

The amplitude of the first formant, A1, and the amplitude of the second formant, A2, were measured at two points in the vowel, since we hypothesized that a reduction in one or both might conceivably alter perception enough to cause a change in labeling. A 15 ms Hanning window was used in Xwaves to generate a FFT spectrum used in amplitude measurements. The first measurement was taken near the middle of the vowel by centering the analysis window on the positive-going portion of the pitch period closest to the vowel midpoint; this gave rise to estimates of A1 and A2 in decibels. The second measurement was taken near the end of
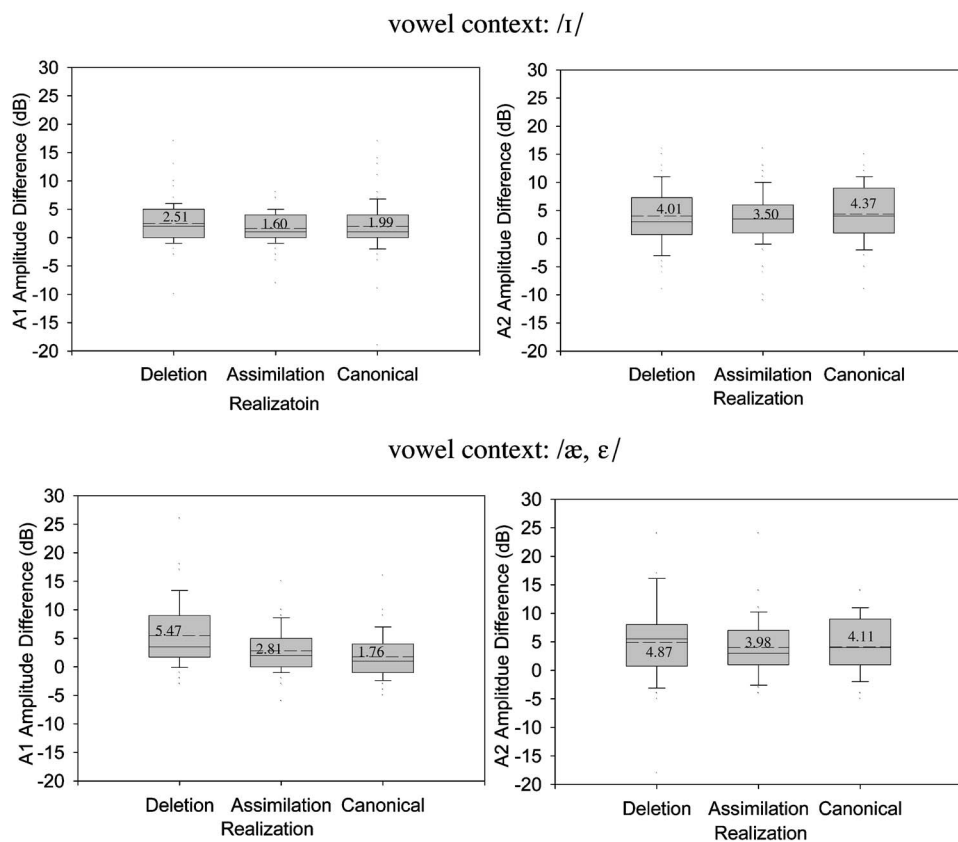
vowel context: /ɪ/

vowel context: /æ, ɛ/

FIG. 5. Box plots of the drop in formant amplitude in tokens labeled as deleted, assimilated, and canonical. The top and bottom graphs differ with respect to the identity of the preceding vowel ([ɪ] or [æ,ɛ]). Data in the left graph are for the first formant, those in the right for the second. The solid line in the middle box is the median of the distribution, and the dashed line is the mean, with its value listed.

the vowel by centering the analysis window on the positive-going portion of the pitch period closest to but not less than 7.5 ms (half the size of the analysis window) from the end of the vowel segment. Two relative amplitude metrics, an A1 difference and an A2 difference, were then calculated by subtracting the values taken at the vowel end point from those taken at the vowel midpoint. No amplitude measurements were taken if the vowel was less than 30 ms in duration. This resulted in discarding eight tokens each from deleted and assimilated categories and seven from the canonical category in the [ɪ] context, as well as seven tokens from the deleted category and two from the assimilated category in the [æ] and [ɛ] contexts.

The results of the amplitude analysis are shown in Fig. 5. A three-way ANOVA was performed on the amplitude measurements, with vowel context and realization as between-item factors and formant amplitude as a within-item factor. The main effect of realization was marginally reliable, $F(2,332)=2.967$, $p<0.053$. None of the post hoc comparisons between pairs of conditions reached significance. Nevertheless, across all four graphs in Fig. 5 there is a trend in the predicted direction, with the drop in energy being greater for segments labeled as deleted than assimilated.

However, additional properties of the graphs do not instill confidence in the ability of listeners to discriminate reliably between these two realizations using amplitude drop-off. The bulk of the deleted and assimilated distributions in three of the four graphs (all but the lower left) overlap. Furthermore, these means differ by no more than 1 dB from each other. In fact, one-way ANOVAs on these triplets of distributions produced no reliable effect of token label. Only

in the lower left graph was the ANOVA reliable, $F(2,123)=6.99$, $p<0.001$. Post hoc Tukey's tests confirmed that the 2.6 dB drop from the deleted to the assimilated condition is reliable. Although amplitude drop-off is greater for deletions than assimilations, its small magnitude combined with wide variability make this acoustic property, just like the F2 difference, a minimally informative cue to use in discriminating deletions from assimilations.

Other reliable effects in the omnibus ANOVA included a main effect of formant amplitude, $F(1,332)=26.766$, $p<0.001$, with drops generally being greatest for A1. This variable also interacted with realization, $F(2,332)=3.87$, $p<0.022$, with the A1 vs A2 differences being much larger for canonical than either deleted or assimilated realizations.

D. Closure duration

What other acoustic characteristic could distinguish tokens labeled as deleted from those labeled as being present (i.e., assimilated or canonical)? For tokens which were labeled as assimilated, deleted, and canonical, the acoustic realization of the consonant(s) at the word boundary, namely $(C_1)\#C_2$ in underlying $VC_1\#C_2(V)$ contexts, consistently corresponded to a single low- or zero-amplitude consonantal occlusion, with no release burst for C_1 if present. In other words, to the extent that the initial stop, C_1, might be present, its most salient acoustic hallmark was as an unreleased stop closure, followed immediately by the consonantal closure for C_2, giving rise in most cases to an otherwise undifferentiated silent interval. Previous research has shown that the perceptual salience of stop consonants depends in part on the dura-

tion of the closure, with listeners tending to perceive a stop C_1 as being absent in $VC_1\#C_2V$ contexts when the duration of the stop closure for $C_1\#C_2$ is short (Repp, 1978; Fujimura *et al.*, 1978; Schouten and Pols, 1983; Ohala, 1990). We reasoned, therefore, that a difference in the entire duration of the (silent) consonantal closure might distinguish instances labeled as deleted from those labeled as assimilated or canonical; if a consonant C_1 in $C_1\#C_2$ were produced with a very short overall closure duration, this might tend to make underlying C_1 segments be perceived as absent, and thus labelers would tend to code it as deleted. Conversely, we reasoned that a segment which was produced more carefully would be perceived as being present, and thus labeled as assimilated or canonical and exhibit a longer closure duration.

Using the same subsets of tokens for which amplitude measurements were taken, we compared closure durations associated with consonantal constrictions at the word boundary across instances labeled as deleted, assimilated, or canonical.[6] Xwaves displays of a waveform and wide-band spectrogram were used to identify the start and end of the consonantal constriction in the vicinity of the word boundary for each token. The starting and ending points of the closure were taken as the positions at which the amplitude suddenly dropped off or increased, respectively, across frequencies. For tokens classified as assimilated or canonical, the consonantal constriction included the duration of the word-final assimilated or canonical consonant, plus the closure period associated with the following word-initial consonant before any burst release. For tokens classified as deleted, the consonantal constriction included only the duration of the closure period associated with the following word-initial consonant prior to any burst release, since phonetic labels indicated zero duration for the word-final consonant.

Box plots of the closure durations for the deleted, assimilated, and canonical tokens are graphed in Fig. 6. The data are impressively consistent across the two vowel contexts. Relative to the deleted distribution, the assimilated distribution is shifted upward by an average of 20 ms into regions of longer durations. The canonical distribution is shifted into even longer regions by the same amount. In the [ɪ] context, a one-way ANOVA across realization conditions was reliable, $F(2,231)=18.62$, $p<0.001$. Statistical analyses in the [æ,ɛ] contexts yielded the same outcome, $F(2,219)=19.99$, $p<.001$.[7] Post hoc Tukey's HSD tests in both vowel contexts showed that all conditions were significantly different from one another. When the data are combined into a single two-way analysis with vowel as the additional factor, only an effect of realization emerged, $F(2,360)=33.19$, $p<0.001$.

Rather than being differentiated by amplitude or F2 transitions, assimilations and deletions were distinguished by the duration of consonantal closures. These results suggest that labelers' judgments about the presence of the word-final stop were likely to have been influenced by closure duration, a finding which mirrors results from experiments showing that perception of consonant presence and identity is mediated by closure duration (Repp, 1978; Fujimura *et al.*, 1978; Schouten and Pols, 1983; Ohala, 1990; Esposito and Di Ben-
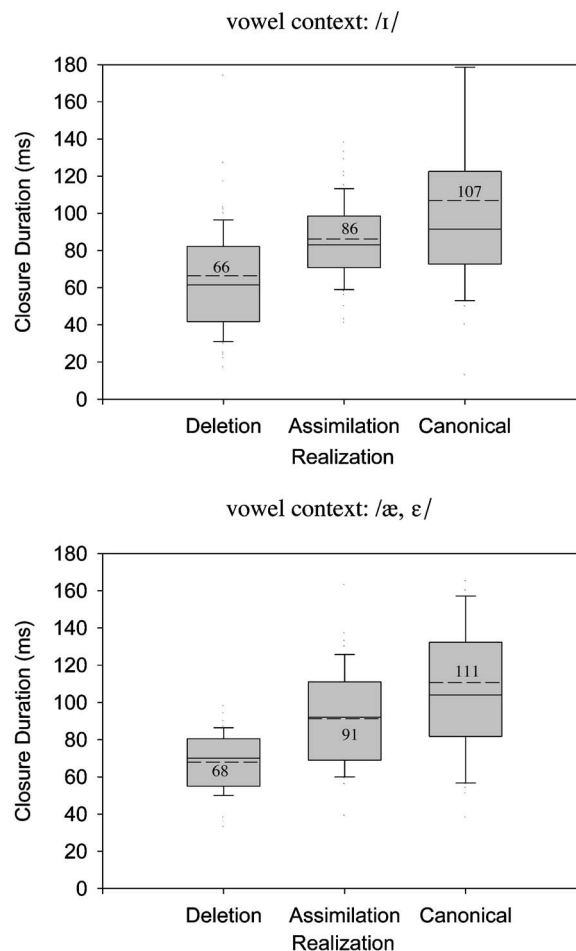
FIG. 6. Box plots of the closure duration in tokens labeled as deleted, assimilated, and canonical (alveolar). The top and bottom graphs differ with respect to the identity of the preceding vowel ([ɪ] or [æ,ɛ]).

etto, 1999). What might be the cause of this significant change in percept? Because gap duration is shorter in the case of segments labeled as deleted, energetic masking of preconstriction material by the postconstriction segment could be greater. However, this explanation seems unlikely given the length of the gap (60 ms; Studdert-Kennedy *et al.*, 1970). A more intriguing possibility is that listeners' sensitivity to the timing in gesture coordination across a word boundary might have caused them to perceive a word-final stop for longer gap intervals (cf. Browman and Goldstein, 1990). Alternatively, a minimum intrinsic duration might be necessary in order for listeners to perceive certain phonemes (Klatt, 1979). In sum, these results indicate that whether an assimilated or deleted segment is perceived is likely to be influenced by closure duration.

IV. GENERAL DISCUSSION

Two studies investigated the nature of regressive place assimilation in a corpus of spontaneous speech. In one study, the frequency of assimilation was examined relative to other kinds of variation, including glottalization and deletion, using phonetic labels from the Buckeye Corpus of spontaneous speech. In a second study, an acoustic investigation was conducted to assess the extent of place modification for tokens

labeled as assimilated versus nonassimilated, as well as the ways in which assimilation differed from other types of variation.

These studies yielded several key findings. First, assimilation is a relatively rare form of pronunciation variation in assimilable environments, as estimated from labeling data produced by trained speech analysts. In particular, assimilation occurred in only 9% of all possible assimilable environments in our data. Tokens labeled as assimilated are likely to correspond to cases which have previously been labeled "extreme assimilations," "complete" or "near-complete assimilations," or "zero alveolars" (e.g., Wright and Kerswill, 1989; Nolan, 1992; Gaskell, 2003). Earlier estimates of rates of these extreme assimilations have varied. For example, Nolan (1992) estimated the prevalence of extreme assimilations to be between 20% and 90% depending on the speed and style of speech. Moreover, Gaskell and Marslen-Wilson (2001) proposed that approximately 50% of assimilations would be "fully assimilated" in speech corpora. In contrast, the present spontaneous speech corpus analysis suggests that the rate of complete or near-complete assimilation in environments where it might occur is low, indicating that the forms which are likely to be most problematic for speech perception are fairly infrequent. In contrast, other types of variation, namely deletion and glottal variants, were much more common in these environments, and canonical realizations also occurred quite often. Lexical frequency was shown to be a factor in the range of variation that a word exhibited: High-frequency lexical items were more likely to appear with multiple variant types than lower-frequency words. Such findings mirror previous results showing that higher frequency lexical items are more likely to exhibit missing segments and other types of modifications (e.g., Bell *et al.*, 2003; Pluymaekers *et al.*, 2005; Ernestus *et al.*, 2006).

Our second finding was that assimilation is a graded phenomenon in spontaneous speech. This can be seen by the fact that alveolars labeled as assimilated showed a range of places of articulation, as gauged by the distribution in F2 values. Moreover, even alveolars labeled as canonical but followed by labial or velar consonants showed a shift in F2 values relative to alveolars that were followed by another alveolar consonant (e.g., bars 4 and 5 in Fig. 3), indicating partial assimilation of the former. These results confirm and extend previous findings using read speech which have shown gradation in acoustic or articulatory markers of degree of assimilation (Gow, 2001, 2002, 2003; Zsiga, 1995; Holst and Nolan, 1995; Nolan, 1992; Wright and Kerswill, 1989). Our results demonstrating gradation in place assimilation are also mirrored by recent results for another conversational speech phenomenon, namely voicing assimilation, which is similarly graded in French as well as Dutch conversational speech (Snoeren *et al.*, 2006; Ernestus *et al.*, 2006).

The finding that assimilation is realized through graded acoustic cues has implications for interpreting the labeling data to estimate rates of assimilation in spontaneous speech and for using such data to study word recognition. Although labelers made categorical labeling judgments about the status of tokens as assimilated versus nonassimilated (e.g., bars 3 and 4 in Fig. 3), the acoustic data showed that tokens in assimilable contexts which were labeled as not assimilated (e.g., *pin point*, bar 4) nevertheless exhibited F2 values which were shifted away from nonassimilable, alveolar contexts (e.g., *did not*, bar 5). This observation suggests that estimates of assimilation rates will depend on labelers' thresholds for hearing tokens as assimilated or not (i.e., as canonical, deleted, or glottal forms). Given that labelers adopted a conservative criterion for the assimilated versus nonassimilated distinction (Kiesling *et al.*, 2006), estimates of assimilation rate based on labeling data provide a reasonable benchmark of near-complete and complete assimilation frequency in spontaneous speech. Finally, these results underscore the necessity of conducting acoustic-phonetic studies to obtain a full picture of connected speech processes, rather than relying on phonological data alone.

A third finding of the present study was that acoustic cues to place of articulation for alveolars labeled as assimilated had virtually identical distributions to those of underlying labial or dorsal sounds. This suggests that assimilation is often complete or nearly complete in spontaneous speech, consistent with findings from read speech (Holst and Nolan, 1995; Nolan *et al.*, 1996; Ellis and Hardcastle, 2002). Together with the finding that assimilation is graded in spontaneous speech, these data suggest that details of assimilation previously examined in read speech material generalize to casual, unscripted speech materials (Wright and Kerswill, 1989; Holst and Nolan, 1995; Zsiga, 1995; Ellis and Hardcastle, 2002).

Our fourth finding concerns acoustic factors which differentiate assimilated, canonical, and deleted variants. Perhaps surprisingly, similar degrees of F2 change were exhibited for assimilated and deleted variants, suggesting that both forms were taking the place of articulation of a following labial or dorsal sound. Moreover, similar relative amplitude levels were found for assimilated and deleted variants, as well as canonical segments (Fig. 5), underscoring the ambiguity of acoustic information in word-final position with respect to the surface realization of a segment. However, the duration of the consonant closure was found to distinguish variant types: both canonical and assimilated variants had longer closure durations than deleted variants. Previous work has shown that listeners are perceptually sensitive to closure duration in judgments related to consonantal context (Esposito and Di Benetto, 1999; Repp, 1978; Fujimura *et al.*, 1978; Schouten and Pols, 1983; Ohala, 1990), lending support that this acoustic cue was likely to be important in distinguishing assimilations from deletions. In sum, these results suggest that the word-final segment is perceptually quite fragile.

The current findings should be qualified in a few respects. First, the present study was limited to talkers from in and around Columbus, OH. While we cannot speculate in depth on how assimilation rate might be different for other dialects, the variety of English spoken in central Ohio is very similar to the General American dialect (Labov *et al.*, 2005). Thus these data provide a reasonable index of assimilation behavior for a large number of speakers in North America. Moreover, with respect to the acoustical study, our investiga-

tion of place assimilation cues was limited to measurements of F2. While information about place of articulation is most readily conveyed both perceptually and acoustically by F2 information in C-to-V and V-to-C transitions (e.g., Liberman *et al.*, 1954; Tartter *et al.*, 1983), F3 transitions, burst spectra, and other cues are also relevant to place perception (e.g., Harris *et al.*, 1958). The present study also did not take prosodic boundaries into account, although likely instances of the largest prosodic boundaries, e.g., full intonational phrase (IP) and utterance boundaries (cf. Beckman and Pierrehumbert, 1986) were excluded from analysis. Prosodic structure has been argued to mediate optional reduction phenomena, such that prosodic phrase-initial positions are realized with markers of articulatory strength, such as glottalization, increased linguopalatal contact, and/or shorter VOT (e.g., Dilley *et al.*, 1996; Fougeron and Keating, 1997). Such cues could be present in an assimilable context, and thus aid processing. However, recent work shows that word-initial alveolar /t/ and /d/ do not show enhanced cues to voicing contrasts in full IP-initial position in American English read speech (Cole *et al.*, 2007), and that prosodic junctures smaller than the full IP may not be important for processing cues to assimilation, at least in Korean (Cho and McQueen, in press). Thus, prosodic structure seems unlikely to have had a significant effect on our results.

What are the implications of these data for theories of spoken word recognition, and in particular, for accounts developed specifically for recognizing assimilated variants (Gaskell and Marslen-Wilson, 1998; Lahiri and Marslen-Wilson, 1991; Gow, 2003)? It seems that they change the nature of the problem that must be solved, in several respects. First, spoken word recognition theories must allow for the fact that there are graded degrees of assimilation in contexts where assimilation is possible, ranging from zero assimilation to complete (i.e., extreme) assimilation. This gradation has yet to be dealt with satisfactorily in most spoken word recognition accounts. However, recent models that have begun to incorporate findings of gradedness in assimilation cues into accounts of how listeners perceive the intended word, such as Gaskell (2003) and Gow (2003), seem the most promising. To facilitate empirical work, future studies might benefit from the use a graded labeling scale (e.g., 1–7) to more explicitly indicate varying degrees of assimilation, as demonstrated in our data.

Second, the present data show that other types of variation in addition to assimilation can occur in assimilable contexts. This suggests that reliance of these models on acoustic cues to resolve the place of articulation of the final segment will need to be revised or augmented. Recognizing a variant pronunciation apparently requires taking into account multiple kinds of acoustic information spread out over time, since place of articulation cues alone are insufficient to distinguish assimilated, deleted, canonical, and glottal variants.

How might spoken word recognition theories account for these other types of variation in assimilable environments, in particular, for deleted and glottal variants? Two possibilities come to mind. Under the view closest to existing accounts of processing of assimilated variants (e.g., Gaskell and Marslen-Wilson, 1998), separate mechanisms

may be involved in recognizing variants which are realized through distinct acoustic dimensions. According to this view, different processes are entailed in recovering segments after modifications to place of articulation, duration, voice quality, and so forth. This is the view implied by traditional linguistic analyses (e.g., Kager, 1999), according to which categorically distinctive phonological processes apply to canonical forms to yield variant types (glottal, deleted, or assimilated) with rather divergent acoustic structures; recovering the intended segment then involves processes of "undoing" individual rule applications to yield the underlying phoneme.

Alternatively, a somewhat different view that was suggested by a reviewer is that assimilated and deleted variants might be dealt with through the same mechanism. This alternative proposal arises from the observation that both assimilated and deleted variants in this study showed comparable degrees of F2 modification, indicating that both could potentially be handled in the same way by the recognition system. Under this proposal, the assimilation rate would be higher, and only glottal variants would require special treatment. However, it does not seem to us that assimilated and deleted variants could be processed in exactly the same way, since the two were distinguished by a timing variable, namely, the duration of a consonantal closure. This suggests that speech timing must be taken into account in spoken word recognition, consistent with recent findings showing that temporary lexical activation of embedded words (e.g., *ham* in *hamster*) depends partly on temporal cues (e.g., Salverda *et al.*, 2003). Moreover, the importance of timing information in speech perception more generally is well established (e.g., Repp *et al.*, 1978; Tartter *et al.*, 1983).

Because current theoretical accounts were not designed to explain the manifold variation in assimilable environments, it might be more appropriate to ask how well-positioned they are to incorporate these new findings. The importance of closure duration leads us to believe that Gow's (2003) account could fare particularly well. This is because the feature parsing account proposed by Gow specifically invokes perceptual principles of auditory scene analysis, an area which focuses on the problem of assigning sounds to their environmental sources. Sounds which are temporally closer tend to preferentially group together into auditory "streams" (e.g., Bregman, 1990), thus potentially providing the rudiments of an account of perception of a deleted segment. Nevertheless, the prevalence and distinct acoustic manifestation of glottal variants pose further challenges for this and other processing accounts, particularly since glottalization may mark vowel onsets and/or prosodic phrase boundaries, in addition to realizing [?] variants of /t/ and /d/ (Dilley *et al.*, 1996; Redi and Shattuck-Hufnagel, 2001).

Given our findings that the final segment is often acoustically unclear, it seems likely that in such cases the surrounding context is the most frequent source of information with which to ensure correct recognition, further increasing the complexity of a theoretical account. One possibility in particular is that the preceding lexical context could aid in interpreting the ambiguous acoustic information. Such a lexically driven restoration process could be highly successful because the segment occurs at the end of the word, where

lexical influences are greatest ([Pitt and Samuel, 1993](#), [1995](#)). However, the success of a lexically guided account depends on the context itself being unique enough to specify how the ambiguous word-final cues should be interpreted. Content words that become lexically unique before word offset would be maximally informative in this regard. Examination of the assimilated and deleted tokens from the acoustic analyses showed that such ideal conditions for restoration occur infrequently. Although word length ranged from 2 to 11 phonemes, 85% were two and three phonemes long, with "it," "in," and "that" being the most frequent. Only 36% of the tokens are lexically unique at or before word offset, but over half of these instances are due to a single word, "that." Removal of this word reduces the number to 19%. These statistics suggest that lexical information will be of only moderate help in recovering the word-final segment. Most likely the larger sentential context will also be needed to determine the word's identity, and thus the identity of the segment ([Gaskell, 2001](#)). Of course, when the acoustic cues unambiguously specify an assimilated segment, lexical and sentential information might not suffice.

V. CONCLUSIONS

In conclusion, the current research shows that regressive place assimilation is only one of several types of variation which occur in assimilable environments, with deletion and glottal variants being others. Assimilation in spontaneous speech was graded, ranging from full to partial to none. Although acoustic cues to place of articulation were consistently present in the signal, they did not determine the type of variant that was heard. Rather, it is necessary to take other kinds of acoustic information into account, such as closure duration, to explain word perception in assimilable contexts.

ACKNOWLEDGMENTS

[1]Consistent with standard linguistic conventions, slashes are used throughout the paper to indicate underlying segment types (e.g., /t/), while square brackets are used to indicate surface phonetic realizations (e.g., a flapped realization, [ɾ]).

[2]Phonetic analysts also selected among labels such as SIL (for silence) or HES (for hesitation) to indicate locations of nonfluent or fluent pauses between portions of running speech which could not be attributed to a stop closure. See Kiesling *et al.* (2006) for more details.

[3]These labeling consistency data compare favorably with other studies (e.g., Irwin, 1970; Eisen, 1991). For example, Eisen (1991) found labeling accuracy of 88% for obstruent consonants.

[4]Glottal variants were not subjected to further acoustic analysis, since their phonetic realizations are quite different from other variant realizations (Redi and Shattuck-Hufnagel, 2001), in a way which did not lend them to comparison along the dimensions of interest. Moreover, critical acoustic analysis for these tokens already took place during the phonetic labeling process, where each token labeled as glottalized was identified as having irregularly timed pitch pulses in the waveform.

[5]For both vowel contexts, 100% of tokens were matched according to the proportion of following labial and velar contexts. Moreover, 94% of tokens preceded by [ɪ] vowels and 91% of tokens preceded by [æ,ɛ] vowels were matched by the gender and age of a talker, while 79% and 66% of these were matched according to the exact identity of a talker, respectively.

[6]It is important to note that the criteria of labelers about the presence and variant status of a particular underlying word-final alveolar token did not expressly take closure duration into account. (See Sec. II A.) Thus, there is no *a priori* reason why one might expect a longer closure duration for assimilated and canonical variants than deleted variants. In fact, the majority of tokens measured here involved underlying stop consonants followed by another stop; the acoustic realization of the closure in such cases was a single, silent interval spanning both segments. Labeling conventions for the Buckeye Corpus (Kiesling *et al.*, 2006) dictated that in such contexts when an underlying stop was judged to be realized as assimilated or canonical, the midpoint of the closure period was taken as the end of the word-final stop, but when an underlying stop was judged to be deleted, the entire closure period was attributed to the following, word-initial stop.

[7]When closure duration is normalized by dividing by the duration of the preceding vowel, which has been proposed as one means of equating for speech rate (e.g., De Jong, 1998), we found that the pattern of results was quite similar.

Barry, M. (**1992**). "Palatalisation, assimilation, and gestural weakening in connected speech," Speech Commun. **11**, 393–400.

Beckman, M. E., and Pierrehumbert, J. B. (**1986**). "Intonational structure in Japanese and English," Phonology Yearbook **3**, 255–309.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., and Gildea, D. (**2003**). "Effects of disfluencies, predictability, and utterance position on word form variation in English conversation," J. Acoust. Soc. Am. **113**, 1001–1024.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).

Browman, C., and Goldstein, L. (**1990**). "Tiers in articulatory phonology, with some implications for casual speech," in *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, edited by J. Kingston and M. E. Beckman (Cambridge University Press, Cambridge), pp. 341–376.

Cho, T., and McQueen, J. "Not all assimilated sounds are perceived equally: Evidence from Korean," J. Phonetics, in press.

CMU pronouncing dictionary (v. 0.6) [www.speech.cs.cmu.edu/cgi-bin/cmudict], Pittsburgh, PA, Carnegie Mellon University (distributor). Date last viewed: 11 June, 2007.

Cole, J., Kim, H., Choi, H., and Hasegawa-Johnson, M. (**2007**). "Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech," J. Phonetics **35**, 180–209.

Connine, C. (**2004**). "It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition," Psychonomic Bulletin and Review **11**, 1084–1089.

Crystal, T. H., and House, A. S. (**1988**). "Segment durations in connected-speech signals: Syllabic stress," J. Acoust. Soc. Am. **83**, 1574–1585.

Dalby, J. M. (**1986**). "Phonetic structure of fast speech in American English," Ph.D. dissertation, Indiana University, Bloomington, IN.

De Jong, K. (**1998**). "Stress-related variation in the articulation of coda alveolar stops: Flapping revisited," J. Phonetics **26**, 283–310.

Dilley, L., Shattuck-Hufnagel, S., and Ostendorf, M. (**1996**). "Glottalization of vowel-initial syllables as a function of prosodic structure," J. Phonetics **24**, 423–444.

Eisen, B. (**1991**). "Reliability of speech segmentation and labeling at different levels of transcription," in *Proceedings of Eurospeech-91*, Berlin, pp. 673–676.

Ellis, L., and Hardcastle, W. J. (**2002**). "Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data," J. Phonetics **30**, 373–396.

Ernestus, M., Lahey, M., Verhees, F., and Baayen, R. H. (**2006**). "Lexical frequency and voice assimilation," J. Acoust. Soc. Am. **120**, 1040–1051.

Esposito, A., and Di Benetto, M. G. (**1999**). "Acoustical and perceptual study of gemination in Italian stops," J. Acoust. Soc. Am. **106**, 2051–2062.

Fougeron, C., and Keating, P.A. (**1997**). "Articulatory strengthening at edges of prosodic domains," J. Acoust. Soc. Am. **101**, 3728–3740.

Fujimura, O., Macchi, M. J., and Streeter, L. A. (**1978**). "Perception of stop consonants with conflicting transitional cues: A cross-linguistic study," Lang Speech **21**, 337–346.

Gaskell, M. G. (**2001**). "Phonological variation and its consequences for the word recognition system," Lang. Cognit. Processes **16**, 723–729.

Gaskell, M. G. (**2003**). "Modelling regressive and progressive effects of assimilation in speech perception," J. Phonetics **31**, 447–463.

Gaskell, M. G., and Marslen-Wilson, W. D. (**1998**). "Mechanisms of phonological inference in speech perception," J. Exp. Psychol. Hum. Percept. Perform. **24**, 380–396.

Gaskell, M. G., and Marslen-Wilson, W. D. (**2001**). "Lexical ambiguity and spoken word recognition: Bridging the gap," J. Mem. Lang. **44**, 325–349.

Gow, D. W. (**2001**). "Assimilation and anticipation in continuous spoken word recognition," J. Mem. Lang. **45**, 133–159.

Gow, D. W. (**2002**). "Does English coronal place assimilation create lexical ambiguity?" J. Exp. Psychol. Hum. Percept. Perform. **28**, 163–179.

Gow, D. W. (**2003**). "Feature parsing: Feature cue mapping in spoken word recognition," Percept. Psychophys.. **65**, 575–590.

Gow, D. W., and McMurray, B. "Word recognition and phonology: The case of English coronal place assimilation," Papers in Laboratory Phonology 9, in press.

Guy, G. R. (**1980**). "Variation in the group and the individual: The case of final stop deletion," in *Locating Language in Time and Space*, edited by W. Labov (Academic, New York).

Harris, K. S., Hoffman, H. S., Liberman, A. M., Delattre, P. C., and Cooper, F. S. (**1958**). "Effect of third-formant transitions on the perception of the voiced stop consonants," J. Acoust. Soc. Am. **30**, 122–126.

Holst, T., and Nolan, F. (**1995**). "The influence of syntactic structure on [s] to [ʃ] assimilation," in *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, edited by B. Connell and A. Arvaniti (Cambridge University Press, Cambridge), pp. 315–333.

Irwin, R. B. (**1970**). "Consistency of judgments of articulatory productions," J. Speech Hear. Res. **13**, 548–555.

Johnson, K. (**2004**). "Massive reduction in conversational American English," in *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*, edited by K. Yoneyama and K. Maekawa (The National International Institute for Japanese Language, Tokyo), pp. 29–54.

Kager, R., (**1999**). *Optimality Theory*. (Cambridge University Press, Cambridge).

Kiesling, S., Dilley, L., and Raymond, W. (**2006**). "The Variation in Conversation (ViC) Project: Creation of the Buckeye Corpus of conversational speech," Department of Psychology, Ohio State University, Columbus, OH, available at www.buckeyecorpus.osu.edu. Last viewed 8/20/07.

Klatt, D. H. (**1979**). "Synthesis by rule of segmental durations in English sentences," in *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Öhmann (Academic, New York), pp. 287–299.

Kohler, K. (**1990**). "Segmental reduction in connected speech in German: Phonological facts and phonetic explanations," in *Proceedings of the NATO Advanced Study Institute on Speech Production and Speech Modelling*, Bonas, France (Kluwer Academic, Dordrecht).

Labov, W., Ash, S., and Boberg, C. (**2005**). *The Atlas of North American English: Phonetics, Phonology and Sound Change* (Mouton de Gruyter, Berlin).

Lahiri, A., and Marslen-Wilson, W. (**1991**). "The mental representation of lexical form: A phonological approach to the recognition lexicon," Cognition **38**, 245–294.

Liberman, A. M., Delattre, P. C., Cooper, F. S., and Gerstman, L. J. (**1954**). "The role of consonant-vowel transitions in the perception of stop and nasal consonants," Psychol. Monogr. **68**, 1–13.

Miller, J. L. (**2001**). "Mapping from acoustic signal to phonetic category: Internal category structure, context effects and speeded categorization," Lang. Cognit. Processes **16**, 683–690.

Mitterer, H., and Ernestus, M. (**2006**). "Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch," J. Phonetics **34**, 73–103.

Neu, H. (**1980**). "Ranking of constraints on /t,d/ deletion in American English: A statistical analysis," in *Locating Language in Time and Space*, edited by W. Labov (Academic, New York), pp. 37–54.

Nolan, F. (**1992**). "The descriptive role of segments: Evidence for assimilation," in *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, edited by D. R. Ladd and G. J. Docherty (Cambridge University Press, Cambridge), pp. 261–280.

Nolan, F., Holst, T., and Kühnert, B. (**1996**). "Modeling [s] to [ʃ] accommodation in English," J. Phonetics **24**, 113–137.

Ohala, J. J. (**1990**). "The phonetics and phonology of aspects of assimilation," in *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, edited by J. Kingston and M. Beckman (Cambridge University Press, Cambridge), pp. 258–275.

Patterson, D. and Connine, C. M. (**2001**). "A corpus analysis of variant frequency in American English flap production," Phonetica **58**, 254–275.

Patterson, D., LoCasto, P. C., and Connine, C. M. (**2003**). "A corpus analysis of schwa vowel deletion frequency in American English," Phonetica **60**, 45–68.

Pitt, M., and Johnson, K. (**2003**). "Using pronunciation data as a starting point data in modeling word recognition," Paper presented at the 15th International Congress of Phonetic Sciences, Barcelona.

Pitt, M., Johnson, K., Hume, E., Kiesling, S., and Raymond, W. (**2005**). "The Buckeye Corpus of conversational speech: Labeling conventions and a test of transcriber reliability," Speech Commun. **45**, 89–95.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., and Fosler-Lussier, E. (**2006**). "Buckeye Corpus of conversational speech" (1st release) [www.buckeyecorpus.osu.edu], Department of Psychology, Ohio State University (distributor), Columbus, OH. Last viewed 8/20/07.

Pitt, M. A., and Samuel, A. G. (**1993**). "An empirical and meta-analytic evaluation of the phoneme identification task," J. Exp. Psychol. Hum. Percept. Perform. **19**, 699–725.

Pitt, M. A., and Samuel, A. G. (**1995**). "Lexical and sublexical feedback in auditory word recognition," Cognit Psychol. **29**, 149–188.

Pluymaekers, M., Ernestus, M., and Baayen, R. H. (**2005**). "Lexical frequency and acoustic reduction in spoken Dutch," J. Acoust. Soc. Am. **18**, 2561–2569.

Raymond, W., Dautricourt, R., and Hume, E. (**2006**). "Word-medial /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors," Lang. Var. Chg. **18**, 55–97.

Redi, L., and Shattuck-Hufnagel, S. (**2001**). "Variation in realization of glottalization in normal speakers," J. Phonetics **29**, 407–429.

Repp, B. (**1978**). "Perceptual integration and differentiation of spectral cues for intervocalic stop consonants," Percept. Psychophys. 24, 471–485.

Repp, B. H., Liberman, A. M., Eccardt, T., and Pesetsky, D. (**1978**). "Perceptual integration of acoustic cues for stop, fricative, and affricate manner," J. Exp. Psychol. Hum. Percept. Perform. **4**, 621–637.

Salverda, A. P., Dahan, D., and McQueen, J. M. (**2003**). "The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension," Cognition **90**, 51–89.

Schouten, M. E. H. and Pols, L. C. W. (**1983**). "Perception of plosive consonants. The relative contributions of bursts and vocalic transitions," in *Sound Structures: Studies for Antonie Cohen*, edited by M. P. R. van den Broecke, V. J. van Heuven, and W. Zonneveld (Foris, Dordrecht), pp. 227–243.

Shockey, L. (**2003**). *Sound Patterns of Spoken English* (Blackwell, Malden, MA).

Snoeren, N. D., Hallé, P. A., and Segui, J. (**2006**). "A voice for the voiceless: Production and perception of assimilated stops in French," J. Phonetics **34**, 241–268.

Stevens, K. (**1998**). *Acoustic Phonetics* (MIT, Cambridge, MA).

Studdert-Kennedy, M., Shankweiler, D. P., and Schulman, S. (**1970**). "Opposed effects of a delayed channel on perception of dichotically and monotically presented CV syllables," J. Acoust. Soc. Am. **48**, 588–602.

Tartter, V. C., Kat, D., Samuel, A. G., and Repp, B. H. (**1983**). "Perception of intervocalic stop consonants: The contributions of closure duration and formant transitions," J. Acoust. Soc. Am. **74**, 715–25.

Wright, S., and Kerswill, P. (**1989**). "Electropalatography in the study of connected speech processes," Clin. Linguist. Phonetics **3**, 49–57.

Zsiga, E. (**1995**). "An acoustic and electropalatographic study of lexical and postlexical palatalization in American English," in *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, edited by B. Connell and A. Arvaniti (Cambridge University Press, Cambridge), pp. 282–283.

When and why listeners disagree in voice quality assessment tasks[a)]

Jody Kreiman,[b)] Bruce R. Gerratt, and Mika Ito
Division of Head and Neck Surgery, UCLA School of Medicine, 31–24 Rehab Center, Los Angeles, California 90095

Modeling sources of listener variability in voice quality assessment is the first step in developing reliable, valid protocols for measuring quality, and provides insight into the reasons that listeners disagree in their quality assessments. This study examined the adequacy of one such model by quantifying the contributions of four factors to interrater variability: instability of listeners' internal standards for different qualities, difficulties isolating individual attributes in voice patterns, scale resolution, and the magnitude of the attribute being measured. One hundred twenty listeners in six experiments assessed vocal quality in tasks that differed in scale resolution, in the presence/absence of comparison stimuli, and in the extent to which the comparison stimuli (if present) matched the target voices. These factors accounted for 84.2% of the variance in the likelihood that listeners would agree exactly in their assessments. Providing listeners with comparison stimuli that matched the target voices doubled the likelihood that they would agree exactly. Listeners also agreed significantly better when assessing quality on continuous versus six-point scales. These results indicate that interrater variability is an issue of task design, not of listener unreliability. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2770547]

I. INTRODUCTION

Previous research has shown that listeners often disagree with one another in their ratings of voice quality (e.g., Cullinan *et al.*, 1963; Ludlow, 1981; Webb *et al.*, 2003; see Kreiman *et al.*, 1993, for review), and suggests that these disagreements are the result of the methods used to gather ratings and not of flawed or inconsistent perceptual abilities (e.g., Gerratt and Kreiman, 2001; Kreiman and Gerratt, 2005). Ideally, measurement protocols for assessing voice quality should control all the known sources of irrelevant or unwanted variability in listeners' responses. For this reason, modeling the different sources of listener variability in voice quality assessment is the first step in developing accurate, reliable, and valid protocols for measuring quality. Such a model could also provide insight into the perceptual processes listeners use when assessing quality. This study examines the adequacy of one model of the sources of variability in listeners' judgments of voice quality, by quantifying the contributions of four factors to the total variance in a set of quality judgments: the instability of listeners' internal standards for different qualities, difficulties isolating individual attributes in complex acoustic voice patterns, measurement scale resolution, and the magnitude of the attribute being measured.

Previous research has documented the importance of these four factors as sources of variability in quality assessments. We have claimed that traditional rating scale protocols require listeners to judge voices by comparing the vocal stimulus to a standard held in memory for the attribute in question. Listeners' internal standards are idiosyncratic, and vary both within and across listeners with a number of factors, including listeners' previous experiences with voices and the context in which judgments are made (Kreiman *et al.*, 1990; Verdonck-de Leeuw, 1998). When listeners are provided with external comparison stimuli as a referent for their judgments, dependence on varying internal standards is eliminated and rater agreement increases (Gerratt *et al.*, 1993; Gerratt and Kreiman, 2001). This comparison stimulus effect forms the first factor in this experiment. The second factor represents difficulty isolating the dimension to be judged in a complex, time-varying vocal pattern. This factor contributes variability in traditional rating scale tasks, and also in tasks where listeners judge quality with respect to comparison stimuli or anchors. In tasks with comparison stimuli, listeners try to compare magnitudes of the quality being studied in the differing acoustic contexts provided by the test and comparison stimuli. When the comparison stimulus is matched to the test stimulus, so that they differ only in the aspect being judged, listener agreement increases, because the dimension of interest moves to the perceptual foreground as listeners compare magnitudes in otherwise matching acoustic patterns. This helps listeners isolate and focus attention on that dimension (Kreiman and Gerratt, 2005). The third factor contributing to variability—scale resolution—measures the error that occurs when listeners have difficulty selecting a response because the quality of a test stimulus falls between two scale values or two comparison stimuli, or when scale steps or comparison stimuli are so close together that listeners cannot discriminate between ad-

---

TABLE I. Experimental Conditions

	Continuous scale	Six-point scale
Comparison stimulus condition (acoustic context)		
Matched exactly to each target voice (*custom* condition)	Experiment 1	Experiment 2
Matched to the target voices with respect to sex only (*generic* condition)	Experiment 3	Experiment 4
No comparison stimulus	Experiment 5	Experiment 6

jacent levels, and consequently must choose between perceptually equivalent response alternatives (Gerratt *et al.*, 1993). The final factor—the magnitude of the attribute being rated—quantifies the systematic variability in listener agreement levels as a function of the mean rating for a voice. This factor incorporates several previous findings relating agreement levels to the characteristics of the specific voices under study. For example, agreement is best near the endpoints of the scale and worse near the midpoint when listeners rate voices on traditional scales for individual qualities (Kreiman and Gerratt, 1998), in part because the acoustic characteristics of some voices correspond to the limiting case of the quality in question (e.g., aphonia is the greatest possible extent of breathiness; Kreiman and Gerratt, 2005) and in part because fewer response alternatives are available at the end of a scale. In contrast, in method of adjustment tasks (in which listeners adjust the parameters of a synthetic stimulus until it matches the target), agreement increases with the mean rating for a voice, presumably reflecting the difficulty of assessing an acoustic characteristic (and the relatively large difference limens) near the threshold of detection (Kreiman and Gerratt, 2005; Shrivastav and Sapienza, 2003; 2006). By modeling these effects, which depend on the specific set of voices under study, this factor serves to tune the model to the individual data set.

Although these previous studies have documented the individual contributions of these factors to rating variability, no model has examined their interactions, quantified the extent to which as a group they adequately account for interrater variability, or determined the size of their relative contributions to overall variability. To this end, six experiments were devised to examine the effects of these factors on listener agreement in judgments of breathiness and noise-to-signal ratios (NSRs) for pathological voices (Table I). We selected these attributes for study because breathiness is often defined in terms of spectral noise levels, and perceived breathiness has been repeatedly associated with the presence of turbulent noise (e.g., Hillenbrand *et al.*, 1994; Klatt and Klatt, 1990; Shrivastav and Sapienza, 2003; Yiu and Ng, 2004). This definitional equivalence between noise levels and breathiness made it reasonable to compare data from experiments in which listeners manipulated an acoustic attribute (the NSR) and experiments in which they rated a perceptual characteristic (breathiness). Listeners are also perceptually sensitive to spectral noise levels (Gerratt and Kreiman, 2001; Kreiman and Gerratt, 2005), so any differences

observed between experimental conditions could unambiguously be assigned to the task, and not to difficulties hearing the quality being measured.

The first experimental factor—the presence or absence of an external comparison stimulus during the voice assessment task—examined the importance of unstable internal standards for different levels of a vocal attribute. In Experiments 1, 2, 3, and 4, listeners were provided with comparison stimuli; in Experiments 5 and 6 they made judgments solely on the basis of their internal standards. The second factor assessed the effect of listeners' (in)abilities to separate the relevant dimension perceptually from a complex pattern, by providing synthesized comparison stimuli that matched the test voices exactly in quality (hereafter referred to as "custom comparison stimuli;" Experiments 1 and 2) or matched them only with respect to the speaker's sex ("generic comparison stimuli;" Experiments 3 and 4). The third factor—scale resolution—was assessed by comparing ratings on responses from a continuous, visual-analog-type scale (Experiments 1, 3, and 5) with a six-point equal-appearing interval-type rating scale (Experiments 2, 4, and 6). The final experimental factor—the distance of a voice from the extreme values of an attribute—was quantified as the overall mean rating for each voice on each scale. This value was included as a covariate in each of these experiments, to assess the contribution of this fourth factor to rating variability. We hypothesized that the highest levels of interrater agreement would occur in Experiment 1 (in which all hypothetical sources of interrater variability were controlled) and the worst agreement in Experiment 6 (nothing controlled).

II. METHOD

A. Voice samples

Forty voices (20 male speakers, 20 female speakers) were selected at random from a library of pathological voices recorded under identical conditions. No attempt was made to select stimuli that possessed any particular quality, nor did we attempt to create a continuum from mild to severe breathiness, although voices did span the range from near-normal to severely deviant in terms of overall level of pathology. Pathological voices are appealing stimuli not only because of their clinical interest, but also because they encompass the full range of quality that can be produced by a human voice. A 1 s sample was excerpted from the middle of a sustained /a/. Vowels were studied (rather than continuous speech) because steady-state vowels are routinely used for evaluating pathological voice quality and carry much information about the voice source; because analysis and synthesis are far more straightforward than for continuous speech; and because the simpler acoustic structure of steady-state vowels typically yields responses from listeners reflecting simpler perceptual strategies that are more easily interpreted.

Each sample was copied using a custom formant synthesizer optimized for precisely modeling pathologic voice quality. Analysis and synthesis procedures are described in detail elsewhere (Kreiman *et al.*, 2006). Briefly, the synthesizer sampling rate was fixed at 10 kHz. Parameters describing the harmonic part of the voice source were estimated by

inverse filtering a representative cycle of phonation for each voice using the method described by Javkin *et al.* (1987). The extracted pulses were least-squares fit with a modified Liljencrants–Fant (LF) source model (Fant *et al.*, 1985; Kreiman *et al.*, 2007), and the LF model parameters were used to specify the harmonic voice source in the synthesizer. The spectral characteristics of the inharmonic part of the source (the noise excitation) were estimated using a cepstral-domain analysis similar to that described by de Krom (1993). Spectrally shaped noise was synthesized by passing white noise through a 100 tap finite impulse response filter fitted to that noise spectrum. To model the F0 contour, F0 was tracked pulse by pulse on the time domain waveform by an automatic algorithm, and a train of LF pulses with the appropriate periods was added to the noise time series to create a complete glottal source waveform. Formant frequencies and bandwidths were estimated using autocorrelation linear predictive coding analysis with a window of 25.6 ms (increased to 51.2 ms when F0 was near or below 100 Hz). The complete synthesized source was filtered through the vocal tract model to generate a preliminary version of the synthetic voice.

B. Listening pretest

Measurement of many acoustic parameters is difficult and often inaccurate when phonation departs from periodicity (e.g., Titze, 1994; Bielamowicz *et al.*, 1996). For this reason, all synthesizer parameters were perceptually adjusted from their measured values by the first author until the synthetic copies perfectly matched the natural target voices. A pretest was used to verify the accuracy of the adjustments and the synthesis. Twelve listeners [UCLA students and staff; 21–55 years of age; mean age=37.9 years; standard deviation (sd)=12.6 years] heard pairs of voices. On half the trials, a synthetic voice sample was paired with its natural counterpart, and on the other half, stimuli were identical. Each pair was repeated twice, for a total of 160 trials/listener. For each trial, listeners were asked to judge whether the two samples were the same or different, and to rate their confidence in their response on a five-point scale ranging from "positive" to "wild guess." Listeners were not allowed to replay the stimuli before responding. Order of voices in "different" pairs was randomized, and the stimulus pairs were rerandomized for each listener. Listeners were tested individually in a double-walled sound suite. Stimuli were presented in free field at a comfortable constant listening level. Testing lasted approximately 30 min.

To provide a measure of the average discriminability of the synthetic and natural tokens, responses were pooled across listeners. Overall rates of correct and incorrect "same" responses (hits and false alarms) were calculated for each voice. Hit rates ranged across voices from 79.2% to 100%, with an average of 95.3% (sd=5.1%). False alarm rates ranged from 41.7% to 100%, with an average of 73.1% (sd=14.5%). Same/different responses for each voice were then combined with confidence ratings to create a ten-point scale ranging from "positive voices are the same" (1) to "positive voices are different" (10). Values of d' were calculated from these data for each voice; values ranged from 0.002 to 0.97 (mean d'=0.43; sd=0.274). Receiver operating characteristics (ROCs) consisting of nine points each were also constructed from these recoded data following the procedure described by Green and Swets (1966; see also Macmillan and Creelman, 2005). The area under the ROC for each voice was calculated, along with 99% confidence intervals around these values. In all cases, these confidence intervals included the chance value of 0.5. These data indicate that listeners were unable to consistently distinguish the synthetic copies from the natural samples. We conclude that the synthetic tokens provide good models of the quality of the natural target voice samples.

C. Comparison stimuli

Four sets of synthetic comparison stimuli were constructed. The first set (*custom comparison stimuli*) comprised 40 synthetic voice tokens modeled on the 40 natural voice samples described earlier. Each of these 40 stimuli was synthesized using the synthesis parameters that produced perceptually exact copies of the voices, as described previously. Listeners manipulated the NSR of these synthetic voices in Experiment 1, as described in the following.

The second set of custom comparison stimuli comprised six additional versions of each synthetic voice, created for use in Experiment 2 (see Table I). The six versions differed only in NSR levels, in steps. All custom comparison stimuli were 1 s in duration. The first stimulus in each of the 40 series was created with NSR equal to −50 dB (noise-free), and the second used the lowest NSR value at which noise was consistently detectable for that voice, as determined by pilot study. This value ranged from −36 to −17 dB, with a mean of −23.8 dB (sd=4.08 dB). The sixth stimulus in each series was created with NSR equal to 0 dB, and the remaining three stimuli evenly spanned the acoustic range between the second and sixth stimuli. The change in NSR between these last four stimuli ranged from 4.25 to 9 dB, with a mean interstimulus step size of 5.94 dB (sd=1.02 dB). Pilot experiments confirmed that all comparison stimuli in each series were easily discriminable from their immediate neighbors in the series.

The third set of comparison stimuli (*generic comparison stimuli*) comprised two synthetic voices, one male and one female. Synthesis parameters for these voices are listed in Table II. Both voices were created using the same LF source pulse shape and F0 contour shape, although mean F0 differed for the male and female tokens (female F0=228 Hz, male F0=135 Hz). Because listeners' sensitivity to changes in the NSR depends, in part, on the shape of the harmonic spectrum (Gerratt and Kreiman, 2001; Kreiman and Gerratt, 2005), these voices were synthesized using a source with a moderate amount of high frequency harmonic energy (neither sinusoidal nor impulse-like; Fig. 1). This resulted in a scale for which small changes in NSR were perceptible, and for which the resulting stimuli sounded natural across NSR values. Vocal tract parameters were modeled on natural male and female voices that were not used elsewhere in these experiments. Listeners manipulated the NSR of these synthetic

TABLE II. Synthesis parameters for generic comparison stimuli (Experiments 3 and 4) [a]

	Female voice	Male voice
Source parameters		
F0 (Hz)	228	135
Tremor rate (Hz)	4	4
Tremor amplitude (Hz)	0.8	0.8
Parameters of the modified LF model:		
t_p	0.002917	0.004063
t_e	0.003866	0.005385
Ee	1.651	1.651
t_2	0.004247	0.005916
t_c	0.004819	0.006712
Vocal tract parameters		
F1/B1 (Hz/Hz)	670/151	792/96
F2/B2 (Hz/Hz)	1128/134	1224/399
F3/B3 (Hz/Hz)	2337/143	2011/557
F4/B4 (Hz/Hz)	3101/263	3027/119
F5/B5 (Hz/Hz)	4200/500	3942/402
NSR values (dB)	−50	−50
	−25	−28
	−18.75	−21
	−12.5	−14
	−6.25	−7
	0	0

[a]These synthetic voices were modeled after natural vowels produced by two individuals. The first two formants of the man's vowel were higher than those for the woman's, but the voices are unambiguously male and female despite this coincidental occurrence.

voices in Experiment 3, as described in the following.

The fourth set of comparison stimuli included six additional versions of these two voices, which were created for use in Experiment 4. As for the custom comparison stimuli, the six versions differed only in NSR levels, in steps. The first stimulus in each six-step series was synthesized with NSR = −50 dB (noise-free). The second stimulus was created with the minimum NSR value at which noise was clearly perceptible in the voice (determined in pilot studies; −25 dB for the female voice, and −28 dB for the male voice). The sixth stimulus had an NSR of 0 dB (extremely noisy). The remaining three stimuli evenly spanned the acoustic range between the second and sixth stimuli. All stimuli were 1 s in duration, and pilot experiments again confirmed that all stimuli in the series were easily discriminable from their neighbors, for both voices.
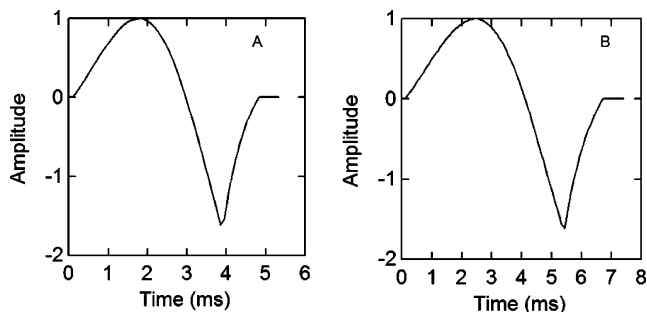

FIG. 1. Source pulses used in synthesis of the generic comparison stimuli in Experiments 3 and 4. (A) Female and (B) male voice sources

D. Listeners and testing environment

A total of 120 listeners (UCLA students and staff members) participated in these experiments, 20 in each task. No listener participated in more than one task, and none had participated in any previous experiments involving these voice stimuli. All listeners reported normal hearing. They ranged in age from 18 to 60 (mean=28.7, sd=10.23).

All six experiments took place in a double-walled sound suite. Stimuli were presented in free field at a comfortable listening level. Listeners practiced with the software using additional voices until they were comfortable and understood the task, but they did not hear the entire stimulus set prior to beginning the experiments. However, recent data (Gerratt and Kreiman, 2007) indicate that listeners who hear voices repeatedly do not adjust the range of their ratings as they learn the stimulus set, but instead remember the score given to each stimulus. This suggests that (lack of) practice with the stimuli did not contribute significantly to the present results. Listeners controlled the pace of the experiments, and were able to play the test and comparison voices as often as they wished, in any order, prior to responding.

E. Experiments 1 and 3: Voice quality assessments using continuous scales and comparison stimuli

In Experiments 1 and 3, listeners heard the 40 test stimuli one at a time in a unique random order. Each natural test voice was paired with a synthetic comparison sample (custom matched to the test voice in Experiment 1, and generic in Experiment 3). For each target voice in each experiment, listeners manipulated the NSR of the comparison stimulus in a method of adjustment task by adjusting a sliding cursor with a mouse until they determined that the synthetic noise level matched that of the natural test voice as closely as possible. The noise scale ranged from −50 dB (noise-free) to 0 dB (extremely noisy), and was set to −50 dB at the beginning of each trial. Listeners were able to play both the test and comparison stimuli in any order as often as they desired as they made their adjustments. In Experiment 1, listeners always judged the test and comparison noise levels in the same acoustic context, because each comparison stimulus matched the corresponding test voice (except for the initial NSR level). Because Experiment 3 used generic comparison stimuli, listeners were required to compare noise levels in unmatched acoustic contexts. Because listeners worked interactively with these voices, no trials were repeated due to the likelihood of remembering individual tokens. Testing lasted about 60 min, on average.

F. Experiments 2 and 4: Voice quality assessments using six point scales and comparison stimuli

Listeners in Experiments 2 and 4 heard the 40 natural test voices one at a time in a unique random order. In each trial, they compared the noise level of the natural test stimulus to those of the tokens in the series of six synthetic comparison stimuli, and decided which of the comparison stimuli best matched the test voice with respect to the amount of noise present. In Experiment 2 listeners compared the test

voices to custom comparison stimuli; Experiment 4 used generic comparison stimuli. In both experiments, listeners were able to replay the test and comparison stimuli as often as they wished, in any order, before responding. Eight randomly selected trials (20%) were repeated to assess test-retest reliability. Testing lasted approximately 25 min on average.

G. Experiments 5 and 6: Voice quality assessments without comparison stimuli

In Experiments 5 and 6, listeners rated the breathiness of each of the 40 test stimuli on traditional rating scales without comparison stimuli. In Experiment 5, they rated the voices by moving a sliding cursor along a line with a mouse (a visual analog-type scale, as in Experiments 1 and 3). Responses were stored as percentages of the scale length from 0 (noise-free) to 100 (extremely noisy). In Experiment 6, listeners rated the breathiness of each stimulus by clicking a number between 1and 6 on a computer display (an equal-appearing interval-type scale, as in Experiments 2 and 4). In both experiments they heard the test stimuli one at a time, in a unique random order. Eight randomly selected stimuli (20%) were repeated to assess test–retest reliability, for a total of 48 trials/listener. Testing took an average of about 15 min to complete.

H. Calculating agreement levels for the different tasks

To assess interlistener agreement, we calculated the likelihood that two randomly selected listeners would agree exactly in their response for a given voice, for each of the 40 voices in each experiment. This approach provides information about systematic patterns of agreement among listeners for specific ranges of rating scales, and contrasts with more commonly used measures like Cronbach's alpha and intraclass correlations that sum across voices to provide a single measure of rater concordance (Kreiman and Gerratt, 1998).

Although no method of calculating agreement can guarantee that listeners mean the same thing when they produce the same response, even within a single protocol, the following procedures were used to equilibrate the response scales in the six experiments as nearly as possible so that agreement levels could be fairly compared across tasks. For tasks using the six-point rating scale, the probability that two listeners would agree exactly was calculated for each voice by examining the difference in ratings between all possible pairs of raters (190 comparisons/voice) and calculating the percentage of exact agreements out of a possible total of 190. For tasks using continuous scales with comparison stimuli, two NSR responses were considered to agree exactly when the difference between them was less than the NSR difference between adjacent comparison stimuli on the six-point scale for that voice. The probability of exact agreement between two raters was calculated based on these quantized ratings in the same manner as it was for the six-point scale protocols. This quantization procedure served to equilibrate the continuous and six-point scales and ensured that responses treated as different were in fact perceptually discriminable. At the same time, this procedure re-centers the window of

TABLE III. Test-retest agreement

	Exact agreement (%)	Mean difference between first and second ratings (sd)
Protocol		
Six-point scale Custom comparison stimuli (Experiment 2)	70.6	0.31 (0.50)
Six-point scale Generic comparison stimuli (Experiment 4)	49.4	0.77 (0.91)
Continuous scale No comparison stimuli (Experiment 5)	66.3	0.52 (0.88)
Six point scale No comparison stimuli (Experiment 6)	50.6	0.68 (0.83)

agreement around each individual response, rather than locking it to the fixed intervals of the six-point scale. Because agreement was not referenced to arbitrary scale divisions, responses that were close together were always considered to agree, regardless of where they fell on the scale, consistent with the continuous nature of these scales.

For ratings on a continuous scale without comparison stimuli (Experiment 5), we defined "exact agreement" as agreement within 20 units on a 100 unit linear scale (five intervals, equivalent to six points). This second quantization procedure had the effect of creating a scale whose intervals are equal in width to those of the six-point scales, but that are again centered around the individual ratings on the continuous scales. As mentioned earlier, responses that were close together were always considered to agree, regardless of where they fell on the scale. The probability of exact agreement was calculated based on these quantized ratings, as described previously.

III. RESULTS

A. Test–retest agreement

Test–retest agreement levels for rating protocols without comparison stimuli (Experiments 5 and 6) and for experiments using six-point rating scales with comparison stimuli (Experiments 2 and 4) are given in Table III. As described previously, test–retest agreement was not assessed for the method-of-adjustment tasks using continuous scales and comparison stimuli (Experiments 1 and 3), due to the probability of learning effects. Listeners were reasonably self-consistent in all these experiments, but for custom comparison stimuli (Experiment 2) the mean difference between the first and second rating was significantly lower than in any other condition [so that ratings agreed more closely; $F(1,636)=9.98$, $p<0.05$; Scheffé post-hoc comparisons, $p<0.05$]. Mean differences between repeated ratings were statistically indistinguishable when listeners made ratings on the six-point scale without comparison stimuli (Experiment 6) and when they made similar ratings with generic comparison stimuli (Experiment 4; $p>0.05$). Test–retest agreement was slightly but significantly higher for ratings on the con-

Kreiman *et al.*: Modeling disagreements in voice quality assessment

TABLE IV. Mean probability of exact agreement between two listeners. Standard deviations are given parenthetically

	Scale resolution	
	Continuous scale	Six-point scale
Comparison stimulus condition (acoustic context)		
Custom comparison stimuli	0.96 (0.12)	0.63 (0.13)
Generic comparison stimuli	0.42 (0.11)	0.28 (0.08)
No comparison stimuli	0.53 (0.14)	0.30 (0.14)

tinuous scale without comparison stimuli than for the six-point scale ratings with generic comparison stimuli ($p < 0.05$).

B. Between-task differences in response variability

Analysis of covariance (ANCOVA) indicated that comparison stimulus condition (custom, generic, and none), scale resolution (continuous versus six point), and the mean rating for each voice all had significant effects on the probability that two listeners would agree exactly. The independent variables and covariate (mean rating) in this experiment together accounted for 84.2% of the variance in agreement levels. The mean probabilities of exact agreement between two raters are given for each task in Table IV.

Across tasks, the probability of agreement covaried significantly with the mean rating [$F(1,233)=55.50$, $p < 0.05$, 4.3% variance accounted for]. Figure 2 shows the probability of exact listener agreement for the six experiments, plotted as a function of the mean rating for each voice in each task. The
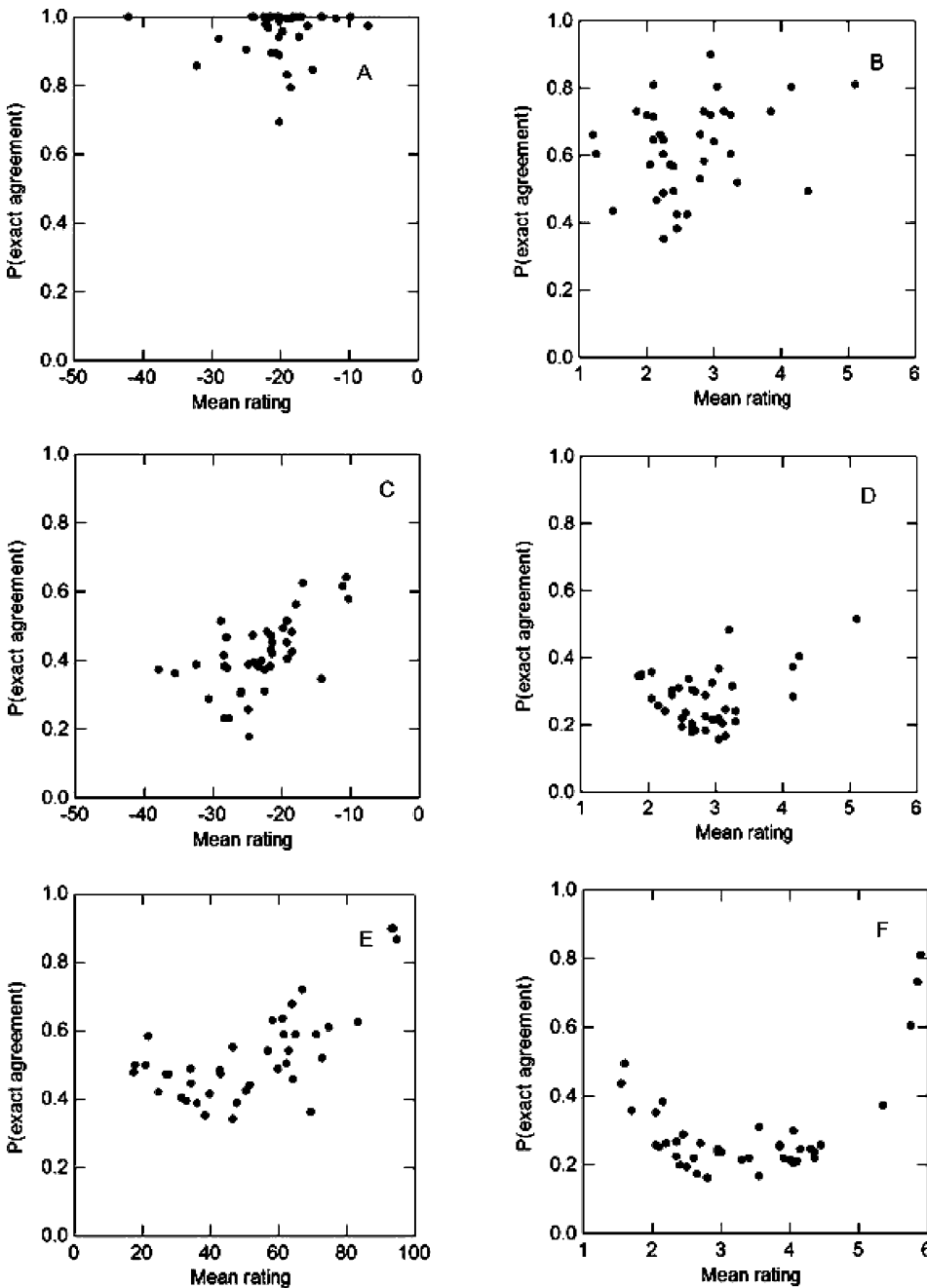
FIG. 2. The probability that two listeners would agree exactly in their rating, versus the mean rating for each voice sample for the six experimental tasks. (A) Continuous scale, custom comparison stimulus; (B) six-point scale, custom comparison stimulus; (C) continuous scale, generic comparison stimulus; (D) six-point scale, generic comparison stimulus; (E) continuous scale, no comparison stimulus; and (F) six-point scale, no comparison stimulus.
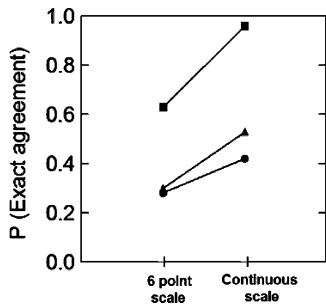
FIG. 3. Interaction between comparison stimulus condition and scale resolution in the ANCOVA analysis. Squares represent custom comparison stimuli; circles represent generic comparison stimuli; and triangles represent the conditions without comparison stimuli.

precise patterns of agreement differed somewhat from task to task. The pattern seen in panels E and F for the traditional rating tasks without comparison stimuli agrees with that usually seen in such rating scale experiments, in which listeners agree best on stimuli at the extremes of the scale and worst in the middle range (Kreiman and Gerratt, 1998). Quadratic equations fit the data in these panels substantially better than did linear models, consistent with this interpretation (panel E: r^2 for the nonlinear model=0.69, versus 0.46 for the linear model; panel F: r^2 for the nonlinear model=0.86, versus 0.14 for the linear model). A similar pattern is apparent in panel D (six-point scale, generic comparison stimuli; r^2 for the nonlinear model=0.41, versus 0.06 for the linear model). In contrast, agreement increased as a linear function of the mean rating for the task using generic comparison stimuli and a continuous scale, although the difference in fit was less pronounced than for panels E and F (panel C; r^2 for the linear model=0.59, versus 0.43 for the nonlinear model). No significant linear or nonlinear relationship was found between the mean rating and agreement levels for the tasks using custom comparison stimuli (panels A and B). (A significance level of $p < 0.05$ was applied for all analyses).

The main effect of comparison stimulus condition accounted for 64.3% of the variance in agreement levels [$F(2,233)=441.58$, $p < 0.05$]. The probability of exact agreement was substantially greater when listeners heard custom comparison stimuli than for the other two conditions [mean=0.79, versus mean=0.35 for generic comparison stimuli, and mean=0.41 when no comparison stimulus was present; planned comparisons, $F(1,233)=651.61$, $p < 0.05$]. The smaller difference between the other two comparison stimu-

lus conditions was also statistically significant: Listeners agreed significantly better when no comparison stimulus was present than they did for generic comparison stimuli [planned comparison, $F(1,233)=4.57$, $p < 0.05$].

A significant main effect of scale resolution was also observed [$F(1,233)=176.73$, $p < 0.05$; 13.0% variance accounted for], with better agreement occurring overall for the continuous scale (mean=0.63) than for the six-point scale (mean=0.40). Differences between scales were significant for all three comparison stimulus conditions, but the advantage provided by a continuous scale was relatively small when generic comparison stimuli were used, leading to a significant interaction effect [Fig. 3; $F(2,233)=115.90$, $p < 0.05$; 2.6% variance accounted for]. This interaction effect appears to have its origins in response biases that inhere in tasks using six point scales with comparison stimuli (Experiments 2 and 4). Listeners in these tasks produced ratings that were significantly lower overall than those in the other four experiments, which did not differ [listener responses standardized to a fraction of the scale length prior to comparison; $F(5,234)=18.11$, $p < 0.05$; Scheffé post-hoc comparisons, $p < 0.05$]. (This effect can be seen in Fig. 2 as a shift to the left for the bulk of the data in panels B and D.) This suggests that when choosing between a comparison stimulus that contains too little noise and one that contains too much noise, listeners consistently selected the less-noisy alternative as the best match, so that mean ratings consistently underestimate the level of noise present in the stimulus relative to the other experimental conditions. This systematic response bias added error to the rating score itself, but at the same time it increased listener agreement by promoting a homogeneous strategy among listeners. In the case of the six-point scale/ generic comparison stimuli task (Experiment 4), this boost in agreement levels has the net effect of minimizing the difference between that task and the condition using a 6 point scale without comparison stimuli (Experiment 6), as shown in Fig. 3.

The model coefficients generated by the ANCOVA allowed us to quantify the specific contribution each model component makes to listener agreement levels. Coefficients for all effects other than the covariate are shown in Table V, and the equation relating the dependent and independent variables is given in the Appendix . For ease of exposition, values in Table V have been converted to percentage change from mean levels in agreement. The constant in the model

TABLE V. The effects of the experimental variables on predicted levels of exact listener agreement. Each entry shows the values added to or subtracted from the model constant (36.44%) to calculate the likelihood of listener agreement for that experimental condition. The first value in each entry is the main effect of scale resolution; the second value is the main effect of comparison stimulus condition; the third value represents the interaction effect; and the fourth value is the sum of the first three. Adding the covariate correction of 0.3149 times the mean rating for that voice to these values produces the predicted agreement level for each voice in the experiment.

	Continuous scale	Six point scale
Comparison stimulus condition		
Custom comparison stimuli	+9.6+28.1+2.68=40.38	−9.6+28.1−2.68=15.82
Generic comparison stimuli	+9.6−16.1−5.35=−11.85	−9.6−16.1+5.35=−20.35
No comparison stimuli	+9.6−12.0+2.67=0.27	−9.6−12.0−2.67=−24.27

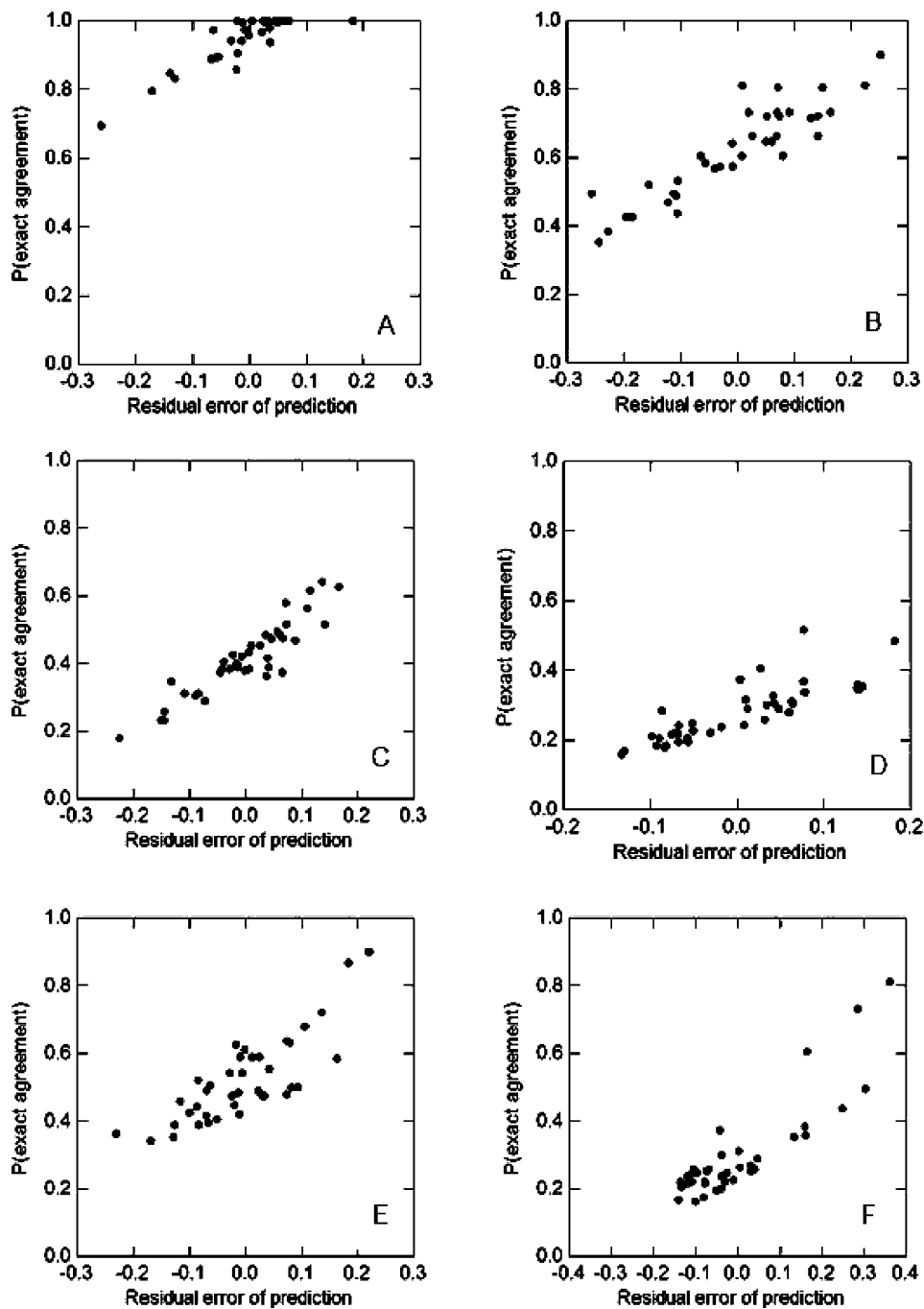FIG. 4. The probability that two listeners would agree exactly in their ratings, versus residual errors of prediction for the six experimental tasks. (A) Continuous scale, custom comparison stimuli; (B) six-point scale, custom comparison stimuli; (C) continuous scale, generic comparison stimuli; (D) six-point scale, generic comparison stimuli; (E) continuous scale, no comparison stimuli; and (F) six-point scale, no comparison stimuli.

equals 36.44%, and represents the average likelihood across all voices and experimental conditions that two listeners will agree exactly. The main effect of comparison stimulus condition produced an increase from this average value of 28.1% in the probability of listener agreement when comparison stimuli matched the test stimuli, a decrease of 16.1% when comparison stimuli did not match test stimuli, and a decrease of 12.0% when no comparison stimuli were available. As a result of the main effect of scale resolution, the probability of exact listener agreement increased from average levels by 9.6% when the scale was continuous, and decreased by the same amount for a 6 point scale. The covariate effect applies equally to values in all cells in the design, and equals 0.3149 times the mean rating (standardized as a fraction of scale length) for that particular voice. The in-

crease in predicted reliability due to this covariate thus ranges from a minimum of 0 at the left end of the scale, to a maximum of 31.49% at the right extreme.

Although overall this model fits the combined data well ($R^2 = 0.842$), fit did vary from task to task. Figure 4 shows for each task the difference between the observed and predicted agreement levels for each voice (residual error) plotted against the probability of exact agreement among raters. For Experiments 1–4 (panels A–D), these plots are unexceptional, with the data more or less symmetrically distributed about 0 error. Different patterns are apparent for the tasks without comparison stimuli (Experiments 5 and 6; panels E and F). For the six-point scale without comparison stimuli (panel F), the pattern of residuals indicates that the statistical model underestimates listener agreement near scale end-

points, yielding positive values of prediction error as indicated by the asymmetry of the data around 0 residual and the bifurcation in the data on the positive end of the scale, corresponding to the two ends of the rating scale in Fig. 2(F). A similar pattern is apparent, although to a lesser extent, in Fig. 4(E), which represents errors of prediction for the task using a continuous scale without comparison stimuli [cf. Fig. 2(E), in which the U pattern is less marked than in Fig. 2(F)]. This pattern of residual error suggests that some of the remaining variance in levels of rater agreement may be due to an interaction between the mean rating and the presence/absence of comparison stimuli in determining the likelihood of listener agreement. To test this hypothesis, we repeated the ANCOVA analysis while omitting the two tasks without comparison stimuli. R^2 increased 4.1% to 88.3% variance accounted for. The effect of the covariate remained statistically significant, but the variance it accounted for decreased from 4.3% to 2.0%.

C. Patterns of agreement across voices

The above results indicate that the context in which a quality judgment is made (represented in this study by the presence and type of comparison stimuli) is the primary determinant of listener agreement levels. This result in turn suggested that in conditions where the comparison stimuli do not match the target voice, the task of assessing quality might be easier in some cases than others (and listeners might agree better in those cases) because of differences in the extent to which the generic comparison stimuli and test stimuli resembled each other, just by chance. To test this hypothesis, we measured the difference between the test stimuli and generic comparison stimuli with respect to the center frequencies of the first three formants, the difference in amplitude between the first and second harmonics (H1–H2), mean F0, and F0 sd, and used these differences to predict the probability of exact listener agreement in the experiments using generic comparison stimuli (Experiments 3 and 4). For the continuous rating scale task, the likelihood of listener agreement increased significantly with similarity in H1–H2 values $[F(1,38)=15.75, p<0.05; r=0.54]$. For the six-point scale task, the likelihood of exact listener agreement increased significantly with similarity in H1–H2 and the center frequency of F3 $[F(2,37)=5.50, p<0.05;$ multiple $R=0.48]$.

IV. DISCUSSION

This study tested the hypothesis that four factors (instability of memory standards for levels of a quality, ability to isolate single dimensions in a complex context, scale resolution, and the magnitude of the attribute being measured) govern the extent to which listeners do or do not agree in voice quality assessments. These four factors accounted for 84.2% of the observed variance in listener agreement levels, with all four factors contributing significantly to this value. Examination of residual errors of prediction (Fig. 4) did not show any other systematic tendencies, suggesting that no other significant factors contribute to predicting agreement levels. This means that a testing protocol that controls the present small set of factors should in fact provide the optimal tool for measuring voice quality, consistent with the theoretical framework described in the introduction. In fact, listeners agreed exactly 96% of the time in the task using custom comparison stimuli and a continuous scale, consistent with this prediction. Although the present study only examined ratings of breathiness and the NSR, we note that breathiness is integrally related to roughness and to overall hoarseness (Kreiman *et al.*, 1994), suggesting that the present results may generalize to ratings of other voice qualities and to the study of voice quality in general.

The numerical model presented in Table V quantifies the contributions of the different experimental factors and their interactions to facilitating listener agreement. Although the precise values associated with each term will probably vary somewhat with the attribute being measured and with the test voices under study, these coefficients can be interpreted as a model of the voice quality perception process. The most important factor in this model is the relative ease of isolating the quality under consideration in the different tasks, corresponding experimentally to provision of custom versus generic comparison stimuli. When comparison stimuli matched the test voices, the likelihood that listeners would agree exactly in their assessment of voice quality more than doubled compared to an unmatched context. When no external comparison stimuli are presented (as in traditional verbal rating scale tasks), listeners must rely on unstable internal standards for vocal qualities and rater agreement levels suffer (as quantified by negative values in Table V for tasks without comparison stimuli); but unstable internal standards are apparently better than mismatched external standards (contrary to our initial hypothesis). At best, providing comparison stimuli that did not match the target voices provided no advantage, and at worst including such stimuli in the assessment protocol actually reduced listener agreement levels relative to all other conditions. These results indicate that listeners can assess a given attribute of a voice pattern most reliably *in the context of the pattern itself*. A matching context facilitates this process; an internally generated standard overall has a negative impact; but providing an unmatched context confuses listeners and is a handicap, not a help, in collecting reliable measurements of voice quality.

Scale resolution also contributes significantly to measurement error: Listeners agreed significantly better on continuous scales than on six-point scales. Further, interactions between scale resolution and comparison stimulus condition explain the fact that agreement levels are similar for Experiments 4 (generic comparison stimuli, six-point scale) and 6 (no comparison stimuli, six-point scale) despite the large negative impact of generic comparison stimuli on agreement levels. As described in Sec. III, response biases were observed when six-point scales were combined with comparison stimuli: Listeners as a group preferred comparison stimuli with too little noise to those with too much noise, with agreement increasing as a result of this consistent response strategy. Simple examination of agreement rates out

of the context of the model in Table V would suggest that test protocols providing generic comparison stimuli are superior to protocols without comparison stimuli, but this impression is a result of the effects of this bias-derived interaction effect, and not of the comparison stimulus condition alone. This observation highlights the difficulty of understanding when and why listeners agree or disagree outside the context of some model of the voice perception process.

The approach taken in this article to the study of listener variability in quality assessment tasks stands in contrast to a recent proposal by Shrivastav and colleagues (Shrivastav *et al.*, 2005), who partitioned rating variance into two components: random errors and systematic biases (e.g., the tendency of one rater to rate all stimuli higher than another rater). They argued that random error can be eliminated post hoc by averaging together many ratings of a given voice from a single rater, after which systematic biases can be corrected by standardizing each listener's mean rating for the voice. In light of the present results, theoretical difficulties are apparent with this approach in addition to the practical limitations noted by the authors. The present results suggest that only a very small portion of rating variance is in fact random (exact agreement equaled 96% in Experiment 1, in which all hypothetical sources of variability were controlled), so that the additional advantage accrued by averaging a large number of ratings is small. Second, normalization of averaged ratings requires the assumption that errors due to response biases remain stable in nature and extent across some temporal testing window, so that there is actually some consistent effect for the normalization to correct. If these biases vary across the window of normalization, then the effect is not consistent, and normalization is inappropriate. Evidence suggests that this assumption is not always correct. Gerratt *et al.* (1993) showed that expert listeners' ratings can "drift" in predictable ways within a single rating session as listeners' unstable internal standards for levels of a quality adapt to the current listening context, so that a voice that sounded moderately rough on first hearing sounds severely rough a few minutes later in the context of many mildly deviant voices. In this case, ratings from one part of a listening session are not directly comparable to those from another part of the session.

The present results also shed light on the variable results produced by previous studies using constant comparison stimuli as perceptual anchors in protocols for assessing the breathiness and/or roughness of pathological voice stimuli (Chan and Yiu, 2002; Gerratt *et al.*, 1993). Such "anchored tasks" are attractive in practical terms, because they are easy to use and therefore potentially applicable in the clinic, but studies have found both increases (Gerratt *et al.*, 1993) and decreases (Chan and Yiu, 2002) in listener reliability relative to traditional unanchored voice quality ratings. In the context of the proposed model of the sources of variability in quality assessments, differences in study design account for these variable findings. Gerratt *et al.* (1993) used test and anchor stimuli drawn from the same synthetic continuum (a custom comparison stimulus task), and found that although agreement among raters improved overall (presumably due to the use of custom comparison stimuli), agreement among listen-

ers decreased the more the target acoustic parameter in the test stimuli differed from that same parameter in the anchor stimuli, an apparent result of the poor resolution provided by their five-point scale. In contrast, listeners in Chan and Yiu (2002) referred to two generic comparison stimuli while performing a visual analog rating scale task. In that study, rating variability for untrained listeners remained unchanged or even increased relative to ratings made without comparison stimuli. This finding is consistent with the present finding that a generic anchor stimulus provides no benefit, presumably because listeners cannot reliably separate the quality being assessed from the overall voice pattern.

Finally, these results suggest that listener perception of voice quality involves a specific combination of holistic, gestalt-like pattern processing and featural analysis. To the extent that listeners rely on individual acoustic dimensions when making quality judgments, their attention to a dimension appears to depend on access to the context of the entire voice pattern. In other words, the features are part of the pattern, and have little or no perceptual importance outside of that pattern. This result is consistent with descriptions of the cognitive and neuropsychological processes underlying recognition of speaker identity from voice. In particular, behavioral data from patients with brain lesions and from control subjects indicate that the importance of individual acoustic cues to speaker identity in voice recognition tasks can be evaluated only in the context of the overall voice pattern (e.g., Van Lancker *et al.*, 1985; Van Lancker and Kreiman, 1987; Neuner and Schweinberger, 2000; cf. neuroimaging results in Belin *et al.*, 2000; Belin *et al.*, 2002; Levy *et al.*, 2001; Warren *et al.*, 2006; see Kreiman and Sidtis, 2008, for extended review). Voice recognition tasks involve assessing voice quality, so this apparent convergence is not surprising, and may point the way for development of a common theoretical framework for understanding the principles underlying many different aspects of voice perception. Studies of voice quality are nearly always specific to some particular discipline or research question (speaker recognition from voice; clinical assessment of voice quality; detection of lying from voice; assessment of the speaker's emotional state; and so on), and no comprehensive model exists to explain how listeners exploit different aspects of the voice signal when performing these various tasks. However, by considering the commonalities between studies of the different uses listeners make of voice quality, it may be possible to develop explanations that unify seemingly disparate areas of study.

APPENDIX: COEFFICIENTS OF THE ANCOVA MODELS

$$A1 = \begin{cases} 1 & \text{if the task uses a six-point rating scale} \\ -1 & \text{when the rating scale is continuous,} \end{cases}$$

$$C1 = \begin{cases} 1 & \text{when the task uses generic comparison stimuli} \\ 0 & \text{when the task uses custom comparison stimuli} \\ -1 & \text{when the task does not include comparison stimuli,} \end{cases}$$

$$C2 = \begin{cases} 0 & \text{if the task uses generic comparison stimuli} \\ 1 & \text{if the task uses custom comparison stimuli} \\ -1 & \text{when the task does not include comparison stimuli,} \end{cases}$$

M = mean rating across listeners for that voice.

When all six experimental tasks are included in the analysis, the probability of exact listener agreement is predicted by

$$\text{Predicted agreement} = 0.3644 + (-0.0963 \times A1) + (-0.1608 \times C1)$$
$$+ (0.2810 \times C2) + (0.0535 \times A1 \times C1) + (-0.0268 \times A1 \times C2) + (0.3149 \times M)$$

Belin, P., Zatorre, R. J., and Ahad, P. (**2002**). "Human temporal-lobe response to vocal sounds," Brain Res. Cognit. Brain Res. **13**, 17–26.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (**2000**). "Voice-selective areas in human auditory cortex," Nature (London) **403**, 309–312.

Bielamowicz, S., Kreiman, J., Gerratt, B. R., Dauer, M. S., and Berke, G. S. (**1996**). "A comparison of voice analysis systems for perturbation measurement," J. Speech Hear. Res. **39**, 126–134.

Chan, K. M. K., and Yiu, E. M-L. (**2002**). "The effect of anchors and training on the reliability of perceptual voice evaluation," J. Speech Lang. Hear. Res. **45**, 111–126.

Cullinan, W. L., Prather, E. M., and Williams, D. E. (**1963**). "Comparison of procedures for scaling severity of stuttering," J. Speech Hear. Res. **6**, 187–194.

de Krom, G. (**1993**). "A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals," J. Speech Hear. Res. **36**, 254–266.

Fant, G., Liljencrants, J., and Lin, Q. (**1985**). "A four-parameter model of glottal flow," STL-QPSR **4**, 1–13.

Gerratt, B. R., and Kreiman, J. (**2001**). "Measuring vocal quality with speech synthesis," J. Acoust. Soc. Am. **110**, 2560–2566.

Gerratt, B. R., and Kreiman, J. (**2007**). "Information conveyed by voices," paper presented at the 153rd Meeting of the Acoustical Society of America, June.

Gerratt, B. R., Kreiman, J., Antoñanzas-Barroso, N., and Berke, G. S. (**1993**). "Comparing internal and external standards in voice quality judgments," J. Speech Hear. Res. **36**, 14–20.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Krieger, Huntington, NY).

Hillenbrand, J., Cleveland, R. A., and Erickson, R. L. (**1994**). "Acoustic correlates of breathy vocal quality," J. Speech Hear. Res. **37**, 769–778.

Javkin, H., Antoñanzas-Barroso, N., and Maddieson, I. (**1987**). "Digital inverse filtering for linguistic research," J. Speech Hear. Res. **30**, 122–129.

Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857.

Kreiman, J., and Gerratt, B. R. (**1998**). "Validity of rating scale measures of voice quality," J. Acoust. Soc. Am. **104**, 1598–1608.

Kreiman, J., and Gerratt, B. R. (**2005**). "Perception of aperiodicity in pathological voice," J. Acoust. Soc. Am. **117**, 2201–2211.

Kreiman, J., Gerratt, B. R., and Antoñanzas-Barroso, N. (**2006**). "Analysis and synthesis of pathological voice quality," Technical report, available at http://www.surgery.medsch.ucla.edu/glottalaffairs/files/GASoftwareManual2006.pdf (last viewed July 11, 2007).

Kreiman, J., Gerratt, B. R., and Antoñanzas-Barroso, N. (**2007**). "Measures of the glottal source spectrum," J. Speech Lang. Hear. Res. **50**, 595–610.

Kreiman, J., Gerratt, B. R., and Berke, G. S. (**1994**). "The multidimensional nature of pathologic vocal quality," J. Acoust. Soc. Am. **96**, 1291–1302.

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., and Berke, G. S. (**1993**). "Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research," J. Speech Hear. Res. **36**, 21–40.

Kreiman, J., Gerratt, B. R., and Precoda, K. (**1990**). "Listener experience and perception of voice quality," J. Speech Hear. Res. **33**, 103–115.

Kreiman, J., and Sidtis, D. (**2008**). *Voices and Listeners* (Blackwell Publishing, Oxford).

Levy, D. A., Granot, R., and Bentin, S. (**2001**). "Processing specificity for human voice stimuli: Electrophysiological evidence," NeuroReport **12**, 2653–2657.

Ludlow, C. L. (**1981**). "Research needs for the assessment of phonatory function," ASHA Reports **11**, 3–8.

Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A User's Guide*, second edition (Cambridge University Press, Cambridge, UK).

Neuner, F., and Schweinberger, S. R. (**2000**). "Neuropsychological impairments in the recognition of faces, voices, and personal names," Brain and Cognition **44**, 342–366.

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.

Shrivastav, R., Sapienza, C., and Nandur, V. (**2005**). "Application of psychometric theory to the measurement of voice quality using rating scales," J. Speech Lang. Hear. Res. **48**, 323–335.

Shrivastav, R., and Sapienza, C., (**2006**). "Some difference limens for the perception of breathiness," J. Acoust. Soc. Am. **120**, 416–423.

Titze, I. R. (**1994**). "Toward standards in acoustic analysis of voice," J. Voice **8**, 1–7.

Van Lancker, D., and Kreiman, J. (**1987**). "Voice discrimination and recognition are separate abilities," Neuropsychologia **25**, 829–834.

Van Lancker, D., Kreiman, J., and Emmorey, K. (**1985**). "Familiar voice recognition: Patterns and parameters. Part I: Recognition of backward voices," J. Phonetics **13**, 19–38.

Verdonck-de Leeuw, I. M. (**1998**). "Perceptual analysis of voice quality: Trained and naive raters, and self-ratings," in *Proceedings of Voicedata98 Symposium on Databases in Voice Quality Research and Education*, edited by G. de Krom, Utrecht Institute of Linguistics, Utrecht, pp. 12–15.

Warren, J. D., Scott, S. K., Price, C. J., and Griffiths, T. D. (**2006**). "Human brain mechanisms for the early analysis of voices," Neuroimage **31**, 1389–1397.

Webb, A. L., Carding, P. N., Deary, I. J., MacKenzie, K., Steen, N., and Wilson, J. A. (**2003**). "The reliability of three perceptual evaluation scales for dysphonia," Eur. Arch. Otorhinolaryngol. **261**, 429–434.

Yiu, E., and Ng, C. Y. (**2004**). "Equal appearing interval and visual analogue scaling of perceptual roughness and breathiness," Clin. Linguist. Phonetics **18**, 211–229.

Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners[a]

Diane Kewley-Port,[b] T. Zachary Burkle,[c] and Jae Hee Lee[d]
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

The purpose of this study was to examine the contribution of information provided by vowels versus consonants to sentence intelligibility in young normal-hearing (YNH) and typical elderly hearing-impaired (EHI) listeners. Sentences were presented in three conditions, unaltered or with either the vowels or the consonants replaced with speech shaped noise. Sentences from male and female talkers in the TIMIT database were selected. Baseline performance was established at a 70 dB SPL level using YNH listeners. Subsequently EHI and YNH participants listened at 95 dB SPL. Participants listened to each sentence twice and were asked to repeat the entire sentence after each presentation. Words were scored correct if identified exactly. Average performance for unaltered sentences was greater than 94%. Overall, EHI listeners performed more poorly than YNH listeners. However, vowel-only sentences were always significantly more intelligible than consonant-only sentences, usually by a ratio of 2:1 across groups. In contrast to written English or words spoken in isolation, these results demonstrated that for spoken sentences, vowels carry more information about sentence intelligibility than consonants for both young normal-hearing and elderly hearing-impaired listeners.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2773986]

I. INTRODUCTION

It is widely believed that speech sounds can be separated into two broad categories, vowels and consonants. Vowels are generally higher in intensity, longer in duration, lower in frequency, and the result of slower movement in the articulators than are consonants. It is often noted that the more rapid change observed for consonant articulation can be considered overlaid on the slower vowel gestures (Perkell *et al.*, 2004). In speech perception, consonants are commonly thought to carry more information about sentence intelligibility than vowels. However, this position was challenged in a preliminary report by Cole *et al.* (1996). Results from their noise-replacement paradigm showed that more information about sentence intelligibility was carried by vowels when only vowel segments were present compared to when only consonant segments were present. In order to examine the role of consonant versus vowel information in spoken sentences by persons with hearing impairment, the present research adapted the noise replacement method described by Cole *et al.* (1996).

Evidence that consonants are processed relatively independently from vowels is found in many areas of research, but is also challenged in others. In seminal work in speech perception, Fletcher (1929) [as discussed in Allen (1996)] argued for independent processing of consonants and vowels. His approach led to the establishment of the articulation index, AI (French and Steinberg, 1947), which today is still one of the better predictors of speech intelligibility performance by persons with hearing impairment. Arguments for the separate decoding of consonants and vowels have been strongly stated by the proponents of the motor theory of speech perception (Liberman *et al.*, 1967; Liberman and Mattingly, 1985). Consonants are stated to be categorically perceived by specialized speech perception processors, while vowels are processed by more general auditory mechanisms. More recently, Caramazza *et al.* (2000) reported evidence that vowels and consonants are processed by distinct neural mechanisms. Caramazza studied errors in speech produced by two Italian-speaking aphasics who had complimentary difficulties producing vowels and consonants and concluded that vowels and consonants have independent status in language production. This independence, however, is clearly challenged by approaches to language based on the syllable as the primary unit of speech (see Studdert-Kennedy, 1998 and Studdert-Kennedy and Goldstein, 2003).

Two recent reports have used experimental techniques to investigate the role of consonant and vowel information in identifying words in sentences by normal-hearing listeners. The first one provided the inspiration for the present investigation. It was a conference proceeding by Cole *et al.* (1996), and will be referred to as Cole96 throughout this report. Cole96 used the TIMIT database and sorted the phonetic segments into three groups of sounds: consonants, vowels, and weaksons. [Weaksons or weak sonorants consisted of liquids, glides, and nasals.] Cole96's first experiment re-

---

placed the information associated with one of the three sound groups with either white noise or a periodic sound composed of sinusoids from 200 to 4000 Hz. Thirty-five high school students listened to altered and unaltered sentences through headphones at a comfortable level. Participants could listen to each sentence up to five times and typed the words they heard into a computer. Almost twice as many words were recognized when vowels (and weaksons) (87.4%) were present opposed to when consonants (and weaksons) (47.9%) were present. There was no effect of whether segments were replaced with noise or a periodic sound.

Two additional experiments were reported in Cole96. Experiment 2 demonstrated that sentence intelligibility was poor (3.1%, chance) when only weak sonorants were present. In experiment 3, Cole96 determined that the location of the segment boundaries in the TIMIT database did not strongly alter the large difference found between vowels and consonants in sentence intelligibility. Cole96 concluded overall that vowels are more important than consonants to the recognition of words in sentences such that there is a 2:1 benefit for vowel-only information compared to consonant-only information.

A second study by Owren and Cardillo (2006) also investigated the role of consonant and vowel information in speech. Methods and results contrasted in several important and interesting ways with Cole96. The hypothesis that Owren and Cardillo explored was different, namely to determine if vowels carry more indexical information about the speaker (gender, identity, age, etc.), while consonants carry linguistic cues that signal meaning. Owren and Cardillo (2006), here after "Owren2006," carefully selected word pairs that differed either by meaning (synonyms or not) or talker (same or different talker within gender). In contrast to Cole96, Owren2006 replaced the vowel and consonant segments, including observable formant transitions, with silence. Also, their task was a same-different discrimination paradigm. Although it is difficult to compare Owren2006 to Cole96, consider the data in Owren2006's experiment 2 where meaning was judged. Owren2006 reported a modest, but significant, improvement in discriminating differences in meaning when consonant-only words were presented compared to vowel-only words. Owren2006 carefully considered many reasons for the fact that their results are opposite to those found in Cole96 (see the following discussion for experiment 1). Owren2006's overall conclusion was that their results are consistent with the hypothesis that the dynamic acoustic cues available in consonants are the ones that signal linguistic meaning in isolated words.

While these recent reports have examined the contributions of vowel and consonantal information in spoken English for normal-hearing listeners, only a few have examined how cochlear hearing impairment may affect these contributions. A study by Owens et al. (1968) compared the performance of hearing-impaired adults listening to sets of words in which responses permitted only vowel confusions or only consonant confusions. They developed their own test with five responses, four likely vowels or "other vowel." The consonant task was the open set W-22 word list (Hirsh et al., 1952) presented at 40 dB SL. Their results indicated vastly more errors for the consonant task (46% correct) than for their own vowel task (94% correct). Thus they concluded that vowel errors were infrequent compared to consonant errors for hearing-impaired listeners. While this is not a direct test of the contribution of vowel versus consonant information to word identification, these results have been taken to support the typical statement that consonants are more important to speech recognition by hearing-impaired listeners than vowels. Apparently such word recognition results have been generalized to overall speech intelligibility, whereas the present experiment will examine intelligibility in sentences.

It is obvious that for a person with a high-frequency sloping audiogram, the audibility of higher-frequency consonant energy is sacrificed, often being completely inaudible. Miller (1951) stated that the weakness in intensity of the consonant sounds is unfortunate because the consonants are more critical for the correct interpretation of speech. This position motivated, in part, a number of studies that manipulated the relative intensity of the consonant to the vowel in short syllables, known as consonant-to-vowel ratio (CVR) studies. Increasing the CVR by increasing consonant intensity has produced some improvement in recognition. However the approach taken by Sammeth et al. (1999) was to maintain the level of three stop consonants and reduce the vowel intensity by 6 or 12 dB. In addition, in some conditions, just the isolated portion of the stop consonant was presented without the vowel, i.e., suggesting an extreme value for the CVR. Two notable results are that identification of the stops by hearing-impaired and normal-hearing listeners was considerably better when the vowel was present than absent, and second, that little difference in performance was obtained between the normal syllables and the ones in which the vowel intensity was reduced. The outcome of this study suggests that the presence of vowels is important for the correct identification of stops in syllables. These results challenge both the independence of vowel and consonant information, and the relative priority that consonants have in the processing of speech by hearing-impaired listeners. However, the outcome of these CVR studies may not be easily generalized to processing of words in sentences.

The purpose of this study was to compare the contribution of information provided by vowels versus consonants to sentence intelligibility when hearing is compromised in the elderly with typical hearing loss. The noise replacement paradigm developed by Cole96 was used in a modified form. In the present experiment, only two groups of sounds were investigated, either consonants or vowels, because weak sonorants were classified as consonant sounds. Sentences were presented in three conditions, with either the vowels or the consonants replaced with speech shaped noise, or unaltered. Two experiments were conducted. The first examined whether the preliminary results of Cole et al. (1996) would be replicated for a group of young normal-hearing listeners at a 70 dB SPL signal level (YNH70). These results provided the baseline data for the main experiment using elderly hearing-impaired listeners and a higher signal level, 95 dB SPL, to assure audibility of the sentences. Thus, in experiment 2 the differences between two groups of listeners, elderly hearing-impaired (EHI95) and young normal-hearing

(YNH95), in identifying words in altered and unaltered sentences were compared. It was hypothesized that: (a) both elderly hearing-impaired listeners and young normal-hearing listeners would perform better with unaltered sentences compared to sentences with either vowels or consonants replaced by noise; (b) the elderly hearing-impaired listeners would have lower performance than young normal-hearing listeners; and (c) young normal-hearing listeners would identify more words correctly for vowel-only compared to consonant-only sentences.

II. GENERAL METHODS

A. Overview of experimental design

The contribution of vowel versus consonant information was investigated using a noise replacement paradigm with sentences from the TIMIT database (Zue *et al.*, 1990; Garfolo *et al.*, 1990). Three listener groups were used. Two groups consisted of young normal-hearing participants and the other group of elderly hearing-impaired participants. In experiment 1 the signal level was set close to a conversation level at 70 dB SPL in order to replicate the Cole *et al.* (1996) study. The signal level for experiment 2 was set to a high level (95 dB SPL) so that the sentences would be reasonably audible for hearing-impaired listeners. Criteria for thresholds for the hearing-impaired listeners were selected to make audibility between vowels and consonants somewhat even over the 4000 Hz bandwidth. Three phonetic conditions were tested: (a) unaltered TIMIT sentences (Full); (b) sentences in which all of the vowels were replaced by noise (i.e., only consonants or Cin); and (c) sentences in which all of the consonants were replaced by noise (i.e., only vowels or Vin). Participants identified the words in the sentences during one test session $1-1/2$ to 2 h in length.

B. Participants

All participants were American English native speakers from the North Midland dialect region (DR3) (Garofolo *et al.*, 1990), including Indianapolis and north. A speaker's dialect region was defined as the geographical area of the United States where he or she lived during their childhood years (age 2–10).

Two young normal-hearing groups ($N=8$, $N=16$) served as listeners in this study. They were all students at Indiana University, ranging in age from 20 to 35 years of age, and were paid to participate. Participants had pure-tone thresholds of 20 dB HL or better, at octave intervals from 250 through 8000 Hz, and normal tympanometric tracings. The right ear was used as the test ear.

Elderly hearing-impaired participants ($N=16$) were recruited through the Indiana University Hearing Clinic. Their ages ranged from 65 to 80 years ($M=72.9$ years) and they were paid to participate. Different audiometric criteria were needed to select the elderly hearing-impaired (EHI) listeners such that the vowels and consonants in the sentences would have somewhat balanced audibility. Participants were required to have a long-standing, bilaterally symmetrical, moderate sensorineural hearing loss, believed to be of cochlear origin based on acoustic reflex testing, case history, normal
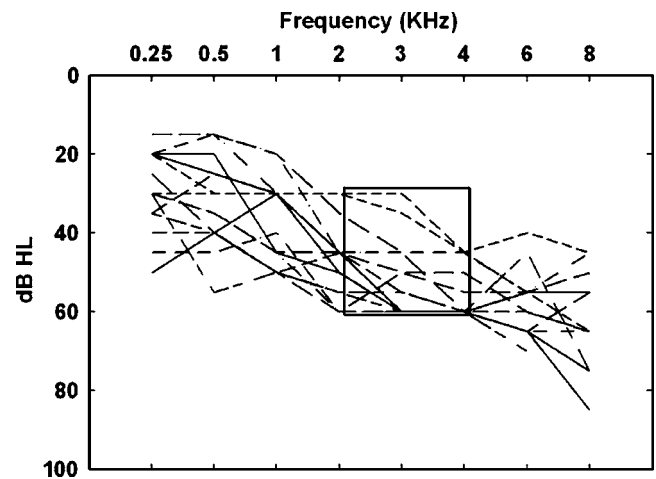

FIG. 1. Hearing thresholds for the 16 subjects in the EHI95 group. The box indicates the hearing criteria used for this study.

tympanograms, and symmetry of air and bone conduction thresholds. Hearing status criteria for subject selection were chosen such that there would be some impairment for vowels at 2000 Hz and good audibility for the consonants between 2000 and 4000 Hz. We selected 35 dB HL as the minimum loss that might affect vowel perception. Criteria for consonants were based on acoustic measurements of the standard long-term speech shaped spectrum (LTSS) (ANSI, 1969). MATLAB was used to generate samples of LTSS noise. Using a calibration vowel and the calibration procedures described in Section E, the 1/3 octave levels of the noise were adjusted to match the low-frequency spectrum of the calibration vowel set to 95 dB SPL. The 1/3 octave band level for the noise at 2000 Hz was 85.6 dB SPL and at 4000 Hz was 67.4 dB SPL. Thus the final criteria for the hearing status of the EHI listeners were thresholds between 35 and 60 dB HL at 2000 and 4000 Hz, yielding audibility for the LTSS of no more than 25 dB at 2000 Hz and no less than 7 dB at 4000 Hz.

The right ear of each EHI participant met all the criteria and was chosen as the test ear. Average thresholds for 0.25, 0.5, 1, 2, and 4 kHz were 29, 32, 37, 48, and 57 dB HL, respectively. Figure 1 shows individual pure-tone thresholds for these 16 EHI participants at octave intervals from 250 through 8000 Hz.

C. Stimuli

The stimuli used in this experiment were extracted from the Texas Instruments/Massachusetts Institute of Technology (TIMIT) corpus (Zue *et al.*, 1990; Garfolo *et al.*, 1990) of read speech. This database contains three types of sentences from 630 speakers representing eight major dialect divisions of American English. We selected the SI sentences from the North Midland dialect region, DR3, that was the best match for listeners from Indiana, Indianapolis, and further north. Forty-eight sentences containing 6–10 words each were chosen randomly with an average of 8.16 words per sentence. For the test sentences, one sentence was picked from each talker, (21 male and female sentences each). The remaining six sentences were used as training/familiarization sentences.

TABLE I. Distribution of sentences over gender and conditions.

	Number of sentences Male	Number of sentences Female	Number of words
Full	7	7	115
Vin	7	7	114
Cin	7	7	114
Total	21	21	343

The 42 test sentences were divided randomly into three test conditions ($N=14$, see Table I). Each condition had about 114 words that averaged four phones/word.

The TIMIT database provides a phonetic classification for American English consonants and vowels. We used this classification with the exception of our following three rules, which were applied in less than 5% of the altered sentences.

(1) Stop closure symbols (i.e., "bcl") were combined with stop (i.e., "b") and treated as a single consonant. This yielded the 32 unique consonants in Table II.
(2) Syllable final V+[r] as in "beer" were transcribed in TIMIT as a vowel plus the consonant [r]. These V+[r] transcriptions were considered as a single rhoticized vowel in this study.
(3) The glottal stop [q] was a separate consonant in TIMIT, generally realized by silence. When the glottal stop was transcribed as occurring between two vowels, [VqV], it was treated as a vowel in this study.

Thus 32 consonants and 20 vowels were used in this experiment as shown in Table II. This is somewhat different than Cole96 in which 12 weak sonorants were placed into a separate category called "weaksons."

D. Sentence processing procedures

In this experiment there were three conditions. "Full" had full phonetic information, i.e., unaltered sentences. "Vin" had the consonants replaced by speech shaped noise. "Cin" had the vowels replaced by speech shaped noise. The digital noise, based on the ANSI (1969) standard idealized long-term average speech spectrum (LTSS), was generated in MAT-LAB. The sample rate for the noise was the same as the TIMIT sentences, 16 000 Hz. The shape of the LTSS signal (flat from 0 to 500 Hz followed by a −9 dB/oct roll-off) was verified by measurement through the ER-3A insert earphones in a 2 cc coupler coupler with the Larson Davis LD2800 spectrum analyzer.

The level of the LTSS replacement noise relative to the overall average level of phonemes in the sentences was set to

TABLE II. Classification of TIMIT phonetic symbols into consonants and vowels.

Group	Phone	Total
Consonant	b d g p t k dx q jh ch s sh z zh f th v dh hh hv l r y w el eng nx m n ng em en	32
Vowel	iy ih eh ey ae aa aw ay ah ao oy ow uh uw ux er ax ix axr ax-h	20

preserve level differences that normally occur for consonants versus vowels. The level differences were measured from eight sentences from the SI group (4 male and 4 female talkers). First, silence was edited out of all sentences. Then using the phonetic boundaries provided in the TIMIT database, the average level of either the consonants or the vowels was calculated. The difference in the average level of consonants relative to vowels was −10 dB. Because the replacement noise was a speech shaped noise, the noise level was referenced to the average level (avl) of all the vowels in the sentences in just the 0–500 Hz passband (approximately 95 dB SPL at the high signal level). Next, informal listening was used to determine that the consonants were not severely masked for various levels of the replacement noise for vowels. For example, equating the noise level to avl was clearly too loud. Eventually, a level of 6 dB less than avl was selected for the vowel replacement noise, and correspondingly the consonant replacement noise level was 16 dB less than avl.

The noise replacement was done using MATLAB scripts using the TIMIT .phn files. When there were adjacent consonants (e.g., CCC as in /str/) or vowels, the noise replaced the entire string. Thus, the duration of consonants and vowels (either singletons or strings) was preserved, and the average replacement noise level was at least 6 dB less than the average levels measured for the vowels or consonants in these sentences. All sentences were normalized to the average rms value measured across all 42 sentences.

E. Calibration

The signal level was set to ensure that no speech segments would be presented at a level greater than 100 dB SPL. To achieve this, first the most intense vowel (MIV_i) in each previously normalized sentence, i, was measured digitally as the rms value for a 50 ms window located around the peak amplitude. Over the 42 sentences, vowels in two sentences had the same highest value of MIV_i (both were highly stressed within their sentences). The vowel /ʌ/ was picked from the word "study" as the calibration vowel. A wave form piece 50 ms long around the peak was duplicated to make a 4 s long calibration signal (calvow4). The rms levels for the other MIV_i values ranged over 11 dB. However, the distribution of the MIV_i showed a narrow peak (mode) such that 71% of the MIV_i values were within ±2 dB of the distribution peak.

All stimuli were played through TDT system II hardware and ER-3A insert earphones at a sample rate of 16 000 Hz. A low-pass filter was designed to reduce the effects of variability in high-frequency EHI audiograms. The filter was flat until 4000 Hz, had the −3 dB cutoff at 4400 Hz, and then a steep slope that fell 50 dB by 5000 Hz (>200 dB/octave). For calibration, the level was measured for the filtered calvow4 vowel played through TDT DA1 16 bit D/A to a HB6 headphone buffer. A 100 dB SPL level was set with ER-3A insert earphones in a 2 cc coupler using the Larson-Davis LD2800 spectrum analyzer and the linear weighting scale. The sentence with the calvow4 (in the word "study") was not too loud in informal listening. With this

TABLE III. Mean and standard deviation (s.d.) of the words identified correctly by each group in each condition are shown in percent correct.

Group	Full		Vin		Cin	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
YNH70	98.58	(2.66)	74.01	(4.39)	33.99	(9.96)
YNH95	99.18	(0.98)	65.06	(5.43)	51.59	(9.77)
EHI95	93.49	(7.75)	40.13	(15.17)	20.02	(16.71)

calibration, the level for the modal peak of the MIV_i vowels was 95 dB SPL; thus, this level is referred to as "95 dB." A second level of "70 dB" was set by attenuating the signal level by 25 dB.

Background noise was used in this study to mask any sentence editing transients. This uniform noise was generated by the TDT WG2 and was also filtered by the 4400 Hz low-pass filter and was continuously present. It was calibrated using the above-mentioned earphones and equipment to have a level less than −50 dB re the 100 dB SPL calvow4.

F. Procedures

Sentences were presented at the 70 or 95 dB level to the right ear of all participants. Listeners practiced with six sentences that were not in the test set, with feedback before beginning the experiment. Half of the participants in each group were assigned to one of two randomizations of the sentence list.

Each sentence was presented twice to the participants at their own pace. Participants were asked to verbally repeat as many of the words as they could identify. Before testing began, simple instructions were given noting that some of the sentences have some noise and sounded a little "choppy." Participants were encouraged to guess. An experimenter sat behind the listeners and scored the responses by hand on a prepared form. A digital recorder was on during the session as backup to assist in scoring. Because the scoring task was quite easy using the form, recordings were rarely consulted.

The procedure to compute a score for each sentence was the same for all listener groups. Words circled on the paper form were correct only if they were spoken entirely correct by the listeners, e.g., no missing suffixes. For each of the two presentations, the experimenter circled the words correct, and any words heard correct in either the first or second presentation were counted. Thus the semantic meaning of the sentences was not considered in this experiment, only correct word recognition.

III. EXPERIMENT 1

The purpose of experiment 1 was to replicate some of the results in the report by Cole96. Cole96's data were obtained at the participants' most comfortable level (MCL). The level we selected for experiment 1 with YNH listeners was 70 dB SPL, a level that is slightly louder than normal conversational loudness but probably similar to one Cole96's listeners selected in the laboratory.

The present experiment attempted to match several experimental variables in the Cole96 study. For example,

Cole96's participants were high school students with no reported hearing problems. As a comparable match, eight young listeners who were students at Indiana University and passed our hearing screening participated in our experiment (called the YNH70 group). Otherwise, the stimuli and procedures for our noise replacement study were designed to be similar to Cole96 (see Sec. II).

A. Results

The young normal-hearing listeners identified words in the Full, Cin, and Vin conditions. Sentence intelligibility was scored as the total number of words correct and converted to percentage values for tables and figures. Intelligibility results for YNH70 listeners are shown in the top row of Table III. As expected, the intelligibility of unaltered sentences (Full) was nearly perfect (98.6%). The percent correct intelligibility for vowel-only and consonant-only sentences was 74.0% and 34.0%, respectively. Clearly the information available in vowels is overwhelmingly superior to that in consonants for the recognition of words in sentences by young normal-hearing listeners.

B. Discussion of experiment 1

Cole96 reported that sentence intelligibility depends more on vowel information than consonant information. Although this experiment differs from that of Cole96 in many details, such as choice of sentences, noise and noise level, and assignment of TIMIT phonetic symbols to consonant or vowel classes, the similarities of the results between the two experiments are striking. In several different noise replacement conditions reported by Cole96, higher scores were observed in the Vin conditions compared to the Cin conditions that typically yielded a benefit ratio about 2:1. In our experiment 1, the ratio of Vin to Cin was 2.2, very similar to Cole96's 2:1 ratio.

A more careful examination of Cole96's paper shows other results relevant to ours. Procedures for the YNH70 group most closely replicated Cole96's experiment 1 when vowels or consonants were replaced by white noise. Interestingly, Cole96's participants performed 13% better in both of their Vin and Cin test conditions than ours. This could be due to the addition of the 12 weak sonorants to both the Vin and Cin scores, although in their experiment 2 word identification from only weak sonorants was just 3%, indicating a small contribution to overall performance. Another reason for the higher Cole96 scores could be due to presentation level differences. Cole96's participants had control over their own presentation level and perhaps the levels they chose

yielded higher performance than our 70 dB SPL (loud conversational speech) level. In spite of higher overall performance in Cole96, the difference between Vin and Cin was 40.0% in our experiment, which was very close to Cole's experiment 1, where the difference between Vin and Cin was 40.8%.

While there is good agreement between our results and those of Cole96, the results of Owren2006 are on the surface quite the opposite: Words with only consonants provide more linguistic information about meaning than words with only vowels. Owren2006 have already provided an excellent discussion of the many reasons why the results between their study with words and ours with sentences are discrepant. (Note that "ours" refers to both the present study and that of Cole96, henceforth to be called Cole/KP.) Four reasons will be discussed here. The Owren2006 stimuli were words with few prosodic cues, while the Cole/KP sentences were rich in prosodic and syntactic cues. Thus listeners could use considerable top-down processing in the sentence intelligibility task, but had to rely on primarily bottom-up acoustic cues to identify just single words. Given that Cole/KP replaced the segments with noise while Owren2006 replaced with silence, perhaps there was some type of masking that interfered with consonant processing in sentences. A slightly different conjecture about the noise replacement paradigm was that the noise might permit phoneme restoration differentially for vowels versus consonants. As noted by Owren2006, these possible reasons for different results between our experiments are testable.

The fourth point of discussion about these experiments deserves more scrutiny: There are many differences in how the boundaries between segments were defined across experiments. In experiments 1 and 2 of Cole96, as well as the present experiment 1, the TIMIT boundaries stored in files in the TIMIT database were used for the noise replacement studies. These boundaries were originally determined by three phoneticians and were intended to assign about half of the formant transitions to the vowels and half to the consonants. Perhaps better performance with vowels as opposed to consonants in word recognition was because the coarticulatory information available in vowels provides enough information about the surrounding consonants to enable the listeners to recognize the intended words. We are examining this in a new study wherein the consonant-vowel boundaries are being systematically shifted. In fact, Cole96 examined the effect of boundary placement in his experiment 3 that included four new experimental conditions such that the vowel or consonant boundaries were either expanded or reduced by 10 ms before replacing the segments with noise. Expanding consonants resulted in a significant improvement in word recognition, whereas expanding vowels improved recognition, but not significantly. Overall, however, the disproportionately high word recognition performance for vowel information over consonant information was maintained.

Experiments in Owren2006 were more explicitly concerned about the role of formant transitions between consonants and vowels. The importance of these "vowel margins" has been well established by Strange and her colleagues

(Strange *et al.*, 1976) in their series of silent-center vowel studies. To reduce the effects of transitions, Owren2006 edited all stimuli such that "visible formant transitions" were deleted between segments. Concerned there might still be dynamic information in the vowel-only segments, in their experiment 3 an additional 50% of the vowel signal was deleted (but consonant-only stimuli remained the same). Results showed a strong reduction in performance for the remaining vowel-only acoustic segments. The conclusions reached by Owren2006 and Cole96 were similar: Formant transitions at the edges of vowels contain more information about the neighboring consonants than the edges of consonants contain about neighboring vowels.

While the importance of the role of formant transitions versus more steady-state acoustic information needs further exploration (see Lee and Kewley-Port, 2006a, b), there are important implications for our experiment 2 that used hearing-impaired listeners in our noise replacement paradigm. The common wisdom in audiology is that the most important segments for word recognition are the consonants, not the vowels. However the present results suggest just the opposite. Namely, vowels may play a more important role for young normal-hearing listeners because they are more intense, longer than consonants, and furthermore contain considerable information about neighboring consonants in their formant transitions. Experiment 2 was designed to explicitly investigate if the role of consonants and vowels in sentence intelligibility would be the same or different for elderly hearing-impaired listeners.

IV. EXPERIMENT 2

A. Methods

Section II described the participants, stimuli, calibration, and procedures used in experiment 2. The criteria for hearing status and sound level used in this experiment were 95 dB SPL such that the vowels and consonants should have been reasonably audible to the elderly hearing-impaired listeners (EHI95). Although this level was somewhat loud for the young listeners (YNH95), this was the level of the loudest vowels in rather short sentences and informal listening indicated that the sound level was not uncomfortable.

In addition to the pure-tone audiometry and tympanometry given to the YNH70 listeners, further clinical evaluations were administered to the 16 EHI95 and 16 YNH95 participants in this experiment. Because the experiment involved repeating sentences, it was necessary to assess participants for possible memory deficits. The Mini-Mental Status Examination (MMSE) (Folstein *et al.*, 1975) and the forward and backward digit span tests were administered. The EHI95 and YNH95 participants all passed the MMSE, and the forward and backward digit span test.

In order to compare results of our experiment with common speech-recognition instruments used in clinical settings, the following speech recognition measures were given in the test ear: speech reception thresholds (SRT), word recognition scores (WRS), and the speech in noise test (SPIN, Bilger *et al.*, 1984). The SRT is a measure of sensitivity (threshold) for spondaic words (ANSI, 1996). Threshold in this study is

defined as the intensity required for 50% recognition of the spondaic words. Word-recognition abilities were assessed in quiet using CID W-22 full lists (50 words) at a presentation level of 40 dB SL [re: speech reception threshold (SRT)]. The SPIN sentences contained 25 high predictability and 25 low predictability sentences that were presented at a level of 95 dB SPL with a signal to babble ratio of +8 dB.

B. Results

1. General results

Average recognition performance for Full, Vin, and Cin sentence conditions for the listener groups are presented in Table III. Consistency of performance was assessed by considering the rank ordering of intelligibility scores across the three tasks. Spearman's rho coefficients were calculated from the YNH95 and EHI95 data that each had 16 listeners, but not the YIH70, which had too few listeners ($N=8$). As expected, rank order coefficients across tasks were small ($r < 0.3$) and not significant for YNH95 listeners because they exhibited little variability as a group. However, the rank order between tasks was moderate ($0.59 < r < 0.69$) and significant for EHI95 listeners indicating reasonable consistent performance within listeners.

Overall, as expected, both the YNH95 and EHI95 listeners performed significantly better in the Full condition compared to Vin and Cin. Clearly, listener groups had no difficulty identifying words in the unaltered sentences at 95 dB SPL. Only one of the participants in the EHI95 group scored poorly on the Full condition, 67.8% correct, whereas the range for the other 15 EHI95 listeners was 88.6%–99.1% correct such that there was considerable overlap with the YNH95 group (range 97.4%–100%). To test if sentence intelligibility in the Full condition was the same for both groups, first a two-way analysis of variance (ANOVA) with two listener groups and three sentence conditions as the repeated measures factor and number of words correct as the dependent variable was implemented. Given the expected significant group × condition interaction ($F(2,60)=23.84$, $p < 0.001$), the Tukey HSD post hoc analysis was applied and showed that for the Full condition, YNH95 and EHI95 listeners performed the same ($p < 0.91$). Similar results were obtained in a separate ANOVA when the number of words correct was transformed to rationalized arcsine units (RAU) to stabilize the error variance (Studebaker, 1985). The fact that these two groups scored equally well for the Full condition verified that the sentences were reasonably audible for the EHI95 group and the listening/repeating task was not difficult to perform. The Full condition was not considered in further analyses.

2. Comparison of YNH95 and EHI95 listeners

Performance by the YNH95 and EHI95 groups to identify the number of words in sentences (NW) in the vowel-only (Vin) and consonant-only (Cin) conditions is shown in Fig. 2 in percent correct. As expected, overall the YNH95 group ($M=58.4\%$) performed significantly better than the EHI95 group ($M=30.0\%$) according to a two-way analysis of variance with two listener groups and two conditions as
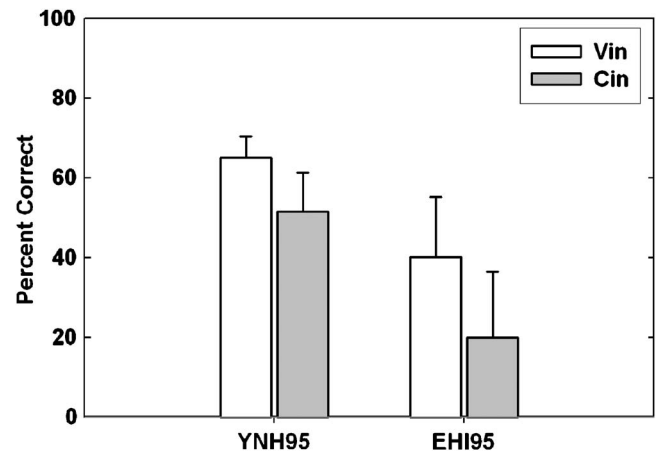

FIG. 2. Percentage of words scored correct in Vin and Cin conditions from YNH95 and EHI95 listeners in experiment 2. Error bars indicate standard deviation.

the repeated measures factor with NW as the dependent measure [$F(1,1)=52.9$, $p < 0.001$]. Significantly better performance was obtained for the Vin condition ($M=52.6\%$) than the Cin condition ($M=35.6\%$) [$F(1,30)=62.1$, $p < 0.001$]. There was a significant decrease in performance for both groups from Vin to Cin conditions with no significant interaction [$F(1,30)=2.45$, $p=0.128$]. According to the Tukey HSD post hoc analysis ($p < 0.001$), performance in the Vin condition (65.1%) was significantly higher in the Cin condition (51.6%) for the YNH95 group. Also, in the EHI95 group, percent-correct performance in the Vin condition (40.1%) was significantly higher than in the Cin condition (19.9%).

C. Discussion of experiment 2

All three of our hypotheses were confirmed for experiment 2. First, both YNH95 and EHI95 listeners recognized more words in the unaltered sentences than the sentences with vowels or consonants replaced by noise. Second, YNH95 listeners performed significantly better than the EHI95 listeners, on average by 28%. As Fig. 2 clearly shows both YNH95 and EHI95 listeners found the Cin task more difficult than Vin. Even though the EHI95 group's average Cin performance was 20%, this is still well above chance given that word identification in sentences is an open response task. EHI95 participants had a wide distribution of scores for the Cin condition (0%–54.4%), and four scored less than 5% correct. Our third hypothesis was also confirmed, namely that the YNH95 group would perform better when vowel-only information was available compared to consonant-only information. Although no specific prediction about hearing-impaired listeners' performance was made, the 2:1 advantage of Vin over Cin conditions for the EHI95 listeners was unexpectedly similar to the typical 2:1 ratio found for normal-hearing listeners in Cole96 and in experiment 1.

The 2:1 ratio was not obtained, however, for the YNH95 listeners at 95 dB SPL where the ratio was 1.26:1. To examine this change in Vin advantage, a comparison of the YNH data for experiments 1 and 2 was made. In experiment 1 the signal level was set near conversational level (70 dB SPL)
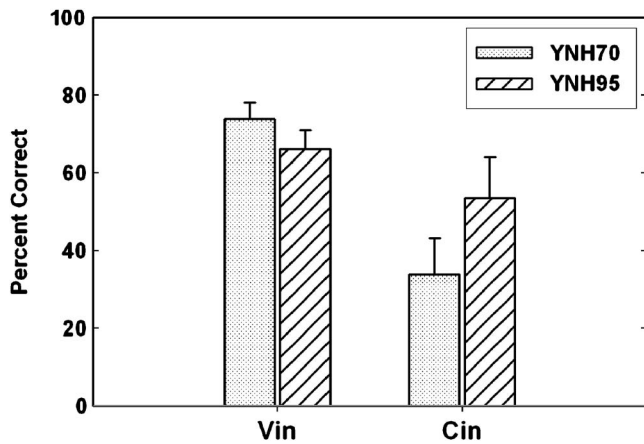
FIG. 3. Percentage of words scored correct in Vin and Cin conditions from YNH70 in experiment 1 and YNH95 in experiment 2. Error bars indicate standard deviation.

for the young normal-hearing listeners to roughly match Cole96's procedures wherein participants set their own most comfortable level (MCL). The higher signal level used in experiment 2 was set to 95 dB SPL to assure audibility for EHI listeners even though it may have introduced some distortions in auditory processing of complex speech sounds (Studebaker *et al.*, 1999). To examine possible negative effects of the high signal level on the YNH95 group ($N=16$), a comparison was made with data from the YNH70 group ($N=8$) in experiment 1. Specifically, data from the YNH70 group ($N=8$) and the first eight participants in the YNH95 group were compared (as shown in Fig. 3). According to a two-way analysis of variance with two listener groups (YNH70, YNH95) and two conditions (Cin, Vin) as the repeated measures factor, overall performance was similar between groups $[F(1,14)=3.15, p=0.10]$ ($M=54\%$ at 70 dB, $M=58.4\%$ at 95 dB) for different presentation levels. However, both condition $[F(1,14)=111.1, p<0.001]$ and the interaction $[F(1,14)=3.15, p<0.097]$ were significant. According to the Tukey HSD post hoc analyses, there was no significant difference in performance (see Fig. 3) between groups in the Vin condition ($p=0.38$). However, the YNH95 group performed significantly better (51.6%) in the Cin condition compared with the YNH70 group (34%) ($p<0.001$).

These results were unexpected. Several recent studies of the "rollover" effects of high presentation levels have shown a decrease in speech intelligibility at higher signal levels (Studebaker *et al.*, 1999; Molis and Summers, 2003; Hornsby *et al.*, 2005). While important differences exist between these studies, a reasonable generalization is that intelligibility for consonants (with higher frequency information) showed greater negative effects of increased level than for vowels where performance was less affected (Molis and Summers, 2003). The present results are consistent with results demonstrating little change for vowels as signal level increases, but are the opposite of the consonant results where the higher signal level increased performance in the Cin condition significantly by 15.6%. The above-cited experiments all used simultaneous noise, whereas the present experiments used replacement noise for the adjacent vowels. Perhaps our procedure allowed the auditory system to get more noise-free

looks at the consonant information that actually resulted in improved performance at the higher sensation level for consonants, although this is very speculative.

Summarizing, for young normal-hearing listeners, a high 95 dB SPL signal level versus a 70 dB SPL level did not change word recognition performance when vowel-only information was present in sentences, and actually improved performance for the consonant-only sentences. This improvement of consonant-only recognition resulted in a reduced vowel advantage of 1.26:1 at the 95 dB SPL level compared to the 70 dB SPL level. We note that broadened auditory filters of EHI listeners may have contributed to their overall reduced performance at 95 dB SPL compared to YNH, but based on results from Richie *et al.* (2005) it is unlikely to account for much of the large 28% difference we observed. The present experiment did not attempt to test EHI listeners at a 70 dB SPL level because the consonants would have had very poor audibility. Thus, it is reasonable to suggest that the vowel advantage at 70 dB SPL for EHI listeners would be extremely high (possibly, greater advantage than 50:1), given that consonant-only recognition was already poor at 20% at 95 dB SPL. Results for EHI95 listeners, however, demonstrated a 2:1 advantage for vowels, indicating a large relative improvement of consonant-only recognition over vowel-only recognition. Heeding the cautionary remarks of Studebaker *et al.* (1999) that audibility and signal level will interact in speech intelligibility, the result that YNH70 and EHI95 listeners have a similar 2:1 vowel advantage in the present noise-replacement task is likely serendipitous. This does not detract, however, from the overall result that listeners with typical hearing impairment will benefit from vowels more than consonants in identifying words in sentences.

D. EHI95 performance in relation to clinical tasks

Standard clinical instruments were used to assess the participant's ability to understand speech in both quiet and in noise in order to relate these clinical scores to scores obtained in the noise replacement experiment. SPIN and WRS were administered to the YNH95 and the EHI95 participants. The SPIN test has three scores, high probability, low probability, and the combination score of both. As expected, the YNH95 group performed better than the EHI95 group on all scores: 96.5% vs 90.8% on the high probability SPIN sentences; 56% vs 34.6% on the low probability sentences, and 76.3% vs 62.7% on the combined scores. For the WRS, the YNH95 group again performed better (98.1%) than the EHI95 group (88.9%).

For the elderly hearing-impaired listeners, the relationship between our novel noise replacement tasks and standard audiologic clinical measures, including audiometric thresholds, was examined using correlational analysis. Because there were many measures near 100% (SPIN High probability sentences and the word scores for Full sentences), all percentage scores were transformed to RAU. First product-moment correlation coefficients were calculated between the

TABLE IV. Correlations between standard clinical instruments and experimental conditions Cin, Vin, and Full ($^*p < 0.05$).

	WRS	SRT	Spin High	Spin Low	Cin	Vin
SRT	−0.10					
Spin High	0.55*	−0.24				
Spin Low	0.42	−0.52*	0.66*			
Cin	0.64*	−0.58*	0.38	0.61*		
Vin	0.74*	−0.33	0.55*	0.66*	0.58*	
Full	0.62*	−0.30	0.51*	0.72*	0.60*	0.64*

Full, Cin, and Vin RAU scores and the clinical scores for the WRS, SRT, and SPIN High and Low probability sentences as shown in Table IV.

As expected, all speech perception measures were positively correlated with one another and negatively correlated with the SRT. The correlation coefficient, r, for $N = 16$ must be greater than 0.49 to be significant at the $p < 0.05$ level. Scores for all three experimental conditions (Full, Vin, and Cin) were significantly correlated at a moderate level with the WRS and with the SPIN low probability score. The full and Vin scores were significantly correlated with the SPIN high probability score, while only the Cin scores were significantly correlated with the SRT.

Correlational analysis was also used to examine the relation among the hearing thresholds in dB HL, age, and performance separately for each of the three experimental conditions. The Cin scores were highly correlated with 2000 and 3000 Hz ($r < −0.85$), while for Vin only moderate correlations ($r < −0.60$) were obtained for 2000 and 3000 Hz. For Full, a moderate correlation of $r = −0.69$ was found for 2000 Hz. No significant correlations ($r < |0.36|$) were found between age and scores for the three experimental conditions. A better characterization of the effects of hearing status for the EHI group was obtained from various pure-tone averages as shown in Table V. The pure-tone averages of 2000, 3000, and 4000 Hz, as well as 3000 and 4000 Hz, were highly correlated, and significant, with the Cin condition. The pure-tone averages of 2000, 3000, and 4000 Hz, as well as 3000 and 4000 Hz, were correlated at a moderate level with the Vin scores.

Because high correlations were observed between the many variables, several linear regression analyses were undertaken to determine how the clinical variables were associated with the scores in our three experimental tasks. Forward, stepwise linear regression in STATISTICA (www.statsoft.com) was selected as the analysis tool. The first linear regression that focused just on audibility was con-

TABLE V. Correlations between pure-tone averages and experimental conditions Cin, Vin, and Full ($^*p < 0.05$).

Pure-tone average	Cin	Vin	Full
500, 1000, 2000	−0.75*	−0.46	−0.42
3000, 4000	−0.81*	−0.57*	−0.43
2000, 3000, 4000	−0.9*	−0.65*	−0.53*

ducted on either the Cin or Vin scores. Only the individual hearing thresholds at 250, 500, 1000, 2000, 3000, and 4000 Hz were used as the variables in these regressions because the sentences had been low-pass filtered at 4000 Hz. Interestingly, the threshold at 2000 Hz accounted for 80% of the variance for the Cin condition, with an additional 13% contributed by thresholds at 3000, 250, and 500 Hz [total 93% of variance, $F(4, 11) = 32.2$, $p < 0.001$]. For the Vin condition, only the thresholds at 1000 and 2000 Hz contributed significantly and accounted for a total of 67% of the variance [$F(5, 10) = 4.17$, $p < 0.026$].

The second regression examined clinical variables more generally for each of the noise replacement conditions. Examining the correlation matrices in Tables IV and V, seven clinical variables were included in the regressions. First the two pure-tone averages that appeared to best represent the speech information, the "PTA Vowel" average of thresholds for 500, 1000, and 2000 Hz, and the "PTA Consonant" average of thresholds 3000 and 4000 Hz were selected. Thus, four clinical tests were included, the measures for the WRS, SRT, SPIN High, and SPIN Low sentences, and finally the elderly listeners' ages.

In the results of the linear regression between the Full RAU score and the seven clinical variables, only the SPIN Low sentence score contributed significantly [$F(1, 14) = 14.7$, $p < 0.002$] with variance accounted for as 51% ($R^2 = 0.51$). Although 12% additional variance was accounted for by WRS, it was not significant. These results are interpreted to mean that for elderly hearing-impaired persons, the identification of words in ordinary sentences, such as the Full sentences, is moderately related to the clinical tasks of scoring one word in a low predictability sentence (SPIN Low), and the identification of single words.

For the linear regression on the Vin RAU scores, only two variables significantly contributed to the regression, WRS with 55% of the variance, and the SPIN Low scores with additional variance 15% [$F(2, 13) = 15.3$, $p < 0.001$]. Notably the WRS and SPIN low scores are the same two variables that were associated with the Full sentence task, although the order of their contributions is reversed. This can be interpreted in a similar way, namely that processing of vowel-only information by elderly hearing-impaired listeners is more related to abilities to processing sound in other speech tasks, such as identification of individual words, whether in lists (WRS) or SPIN Low sentences, than to audiometric and other variables examined.

The linear regression of the Cin scores presented very different results. The variable accounting for 66% of the variance was the PTA for the 3000 and 4000 Hz frequencies. Thus, the poor audibility for the consonantal spectral information was the most important variable contributing to understanding consonant-only sentences. Significant additional variance of 15% was obtained with the PTA of 500, 1000, and 2000 Hz and the SRT in dB (6% variance). Together the performance by elderly hearing-impaired listeners could be predicted from these audibility factors with 87% of the variance, the highest in the regressions for the three experimental conditions.

The results of these regressions suggest that for elderly hearing-impaired listeners, audibility in the higher frequencies is a better predictor of sentence intelligibility from consonant-only information. Of interest is that variability in the pure-tone thresholds of 250, 500, 2000 and 3000 Hz together predicted the Cin scores with very high accuracy, accounting for 93% of the variance. Conversely, the WRS and the SPIN-low variables were better predictors of intelligibility from vowel-only conditions compared to measures of audibility. For the limited age range (65–80 years) spanned by these listeners, aging was not a factor contributing to reduced sentence processing. Naturally the important question of possible effects of cognitive decline with age in sentence processing in this challenging task should be systematically examined in future experiments by using aged-matched listener groups. Because the vowel-only scores were better predicted using speech evaluations, while the consonant-only scores were better predicted by audibility, it can be speculated that vowels and consonants are cognitively processed differently by the auditory system. This is similar to the finding reported by Caramazza et al. (2000) suggesting that vowels and consonants are likely processed by two independent neural mechanisms.

V. SUMMARY AND CONCLUSIONS

This study examined the contribution of information provided by vowels versus consonants to sentence intelligibility in young normal-hearing listeners and in elderly listeners with typical high-frequency hearing loss. A noise replacement paradigm replicated the Cole et al. (1996) finding that for young listeners at conversational levels, there is a 2:1 benefit for sentences that preserve only vowels compared to sentences that preserve only consonants. This same 2:1 benefit for vowel-only sentences was obtained for elderly hearing-impaired listeners in this study at a signal level where our calibration procedures assured audibility of some consonants at frequencies up to 4000 Hz. While three different experiments, Cole et al. (1996), and our experiments 1 and 2 with both young normal-hearing and elderly hearing-impaired listeners, have reported a large benefit for vowel-only relative to consonant-only information, numerous questions about the generality of these results can be raised. We have conducted two additional studies to answer some questions and some results are forecast here. Lee and Kewley-Port (2006a, b) examined issues related to how much information was preserved in the acoustic signal by presenting

portions of the sentences that spanned consonants and vowels (for example, onset transitions from consonants into vowels). It appears that listeners do best when they hear the vowel-only information, and that the acoustic properties of the phonemes (such as stressed versus unstressed vowels) had no effect on the results. It is not possible to speculate, however, if the large vowel benefit found for American English would be obtained for languages with quite a different phonological structure, or for whispered or distorted speech. Fogerty and Kewley-Port (unpublished) moved the consonant-vowel boundaries to extremes and found that formant transitions contribute differently to information available in consonants versus vowels for sentence-recognition tasks. Thus it is quite possible that this series of experiments may have far reaching consequences for theories of speech perception, particularly in relation to our understanding of essential roles that consonant and vowel segments play in speech understanding.

The majority of early research comparing information in vowels versus consonants previously reported focused on CV (consonant-vowel), VC (vowel-consonant), and CVC (consonant-vowel-consonant) monosyllabic words and nonsense syllables. Fletcher's early work (1929) in this area was motivated by the need for a transmission system following World War I. In his experiments, he concentrated on articulation and error patterns made by consonants and vowels. His conclusion that consonants provide more information than vowels for the recognition of isolated words was recently confirmed in Owren and Cardillo (2006). However, we note that linguistic processing of monosyllables relies on segmental, bottom-up information, while our sentence intelligibility task incorporates considerable predictive information from top-down processing. The phoneme error analysis used in Fletcher (1929) was based on discrete symbols to represent the monosyllables. However, incorrectly identified phonemes cannot reveal the actual effect that a specific phoneme has on syllable identification because acoustic information is distributed across phonemes (coarticulation). The noise replacement task removes specific acoustic information and directly measures the accuracy of word identification from the remaining distributed information. Thus our conclusion that vowels carry more information than consonants was assessed directly from acoustic information processed in the auditory system without any specific intervening symbol analysis.

Humans communicate using fluent speech (sentences), and the pitch contours, amplitude envelope, and durational cues in sentences aid in speech intelligibility. Individual words or syllables only contribute to a portion of the intelligibility in a spoken message. In the present research, vowel information was found to have a 2:1 benefit over consonant information in fluent sentences for both young normal-hearing and elderly hearing-impaired listeners. Thus, contrary to popular belief by scientists and clinicians alike, young and elderly listeners performed significantly better when the vowel information was preserved in sentences as opposed to the consonant information. A reasonable implica-

Kewley-Port *et al.*: Contribution of consonant versus vowel information

tion of this result is that design of hearing aids should verify that critical vowel information is preserved by the signal processing algorithms.

Allen, J. B. (**1996**). "Harvey Fletcher's role in the creation of communication acoustics," J. Acoust. Soc. Am. **99**, 1825–1839.

American National Standards Institute. (**1969**). "American National Standards for Calculation of the Articulation Index," *ANSI S3.5, 1969* (American National Standards Institute, New York). NY.

American National Standards Institute. (**1996**). "Specifications for Audiometers," ANSI S3.6, 1996 (American National Standards Institute, New York, NY).

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (**1984**). "Standardization of a test of speech perception in noise," J. Speech Hear. Res. **27**, 32–48.

Caramazza, A., Chialant, D., Capasso, R., and Miceli, G. (**2000**). "Separable processing of consonants and vowels," Nature (London) **403**, 428–430.

Cole, R., Yan, Y., Mak, B., Fanty, M., and Bailey, T. (**1996**). "The contribution of consonants versus vowels to word recognition in fluent speech," Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP'96), Atlanta, GA, May, 1996, pp. 853–856.

Fletcher, H. (**1929**). *Speech and Hearing* (Van Nostrand, New York).

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (**1975**). "Mini-mental state: A practical method for grading the cognitive state of patients for the clinician," J. Psychiatr. Res. **12**, 189–198.

French, N., and Steinberg, J. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.

Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., and Dahlgren, N. (**1990**). "DARPA TIMIT acoustic-phonetic continuous speech corpus CD-ROM," National Institute of Standards and Technology, NTIS Order No. PB91-505065.

Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, G., Eldert, E., and Benson, R. W. (**1952**). "Development of materials for speech audiometry," J. Speech Hear. Disord. **17**, 321–337.

Hornsby, B. W., Trine, T. D., and Ohde, R. N. (**2005**). "The effects of high presentation levels on consonant feature transmission," J. Acoust. Soc. Am. **118**, 1719–1729.

Lee, J. H., and Kewley-Port, D. (**2006a**). "Sentence comprehension when formant transitions are present or absent by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **119**, 3340.

Lee, J. H., and Kewley-Port, D. (**2006b**). "The use of subsegmental information in sentence comprehension with or without formant transitions by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 3347.

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., and Mattingly, I. G. (**1985**). "The motor theory of speech perception revised," Cognition **21**, 1–36.

Miller, G. A. (**1951**). *Language and Communication* (McGraw-Hill, New York).

Molis, M. R., and Summers, V. (**2003**). "Effects of high presentation levels on recognition of low- and high-frequency speech," ARLO **4**, 124–128.

Owens, E., Talbott, C., and Schubert, E. (**1968**). "Vowel discrimination of hearing-impaired listeners," J. Speech Hear. Res. **11**, 648–655.

Owren, M. J., and Cardillo, G. C. (**2006**). "The relative roles of vowels and consonants in discriminating talker identity versus word meaning," J. Acoust. Soc. Am. **119**, 1727–1739.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., and Zandipour, M. (**2004**). "The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts," J. Acoust. Soc. Am. **116**, 2338–2344.

Richie, C., Kewley-Port, D., and Coughlin, M. (**2005**). "Vowel perception in equal audibility conditions by young adults with normal and impaired hearing," J. Acoust. Soc. Am. **118**, 1101–1110.

Sammeth, C. A., Dorman, M. F., and Stearns, C. J. (**1999**). "The role of consonant—vowel intensity ratio in the recognition of voiceless stop consonants by listeners with hearing impairment," J. Speech Lang. Hear. Res. **42**, 42–55.

Strange, W., Verbrugge, R. R., Shankweiler, D. P., and Edman, T. R. (**1976**). "Consonant environment specifies vowel identity," J. Acoust. Soc. Am. **60**, 213–224.

Studdert-Kennedy, M. (**1998**). "The particulate origins of language generativity: from syllable to gesture," in *Approaches to the Evolution of Language*, edited by J. Hurford, M. Studdert-Kennedy, and C. Knight (Cambridge University Press, Cambridge, UK), pp. 202–221.

Studdert-Kennedy, M., and Goldstein, L. (**2003**). "Launching language: The gestural origin of discrete infinity," in *Language Evolution: The States of the Art*, edited by M. H. Christiansen and S. Kirby (Oxford University Press, Oxford), pp. 235–254.

Studebaker, G. A. (**1985**). "A "rationalized" arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (**1999**). "Monosyllabic word recognition at higher-than-normal speech and noise levels," J. Acoust. Soc. Am. **105**, 2431–2444.

Zue, V. W., Seneff, S., and Glass, J. (**1990**). "Speech database development at MIT: TIMIT and beyond," Speech Commun. **9**, 351–356.

Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience

Nathaniel A. Whitmal III,[a] Sarah F. Poissant, Richard L. Freyman, and Karen S. Helfer
Department of Communication Disorders, University of Massachusetts, Amherst, Massachusetts 01003

Channel vocoders using either tone or band-limited noise carriers have been used in experiments to simulate cochlear implant processing in normal-hearing listeners. Previous results from these experiments have suggested that the two vocoder types produce speech of nearly equal intelligibility in quiet conditions. The purpose of this study was to further compare the performance of tone and noise-band vocoders in both quiet and noisy listening conditions. In each of four experiments, normal-hearing subjects were better able to identify tone-vocoded sentences and vowel-consonant-vowel syllables than noise-vocoded sentences and syllables, both in quiet and in the presence of either speech-spectrum noise or two-talker babble. An analysis of consonant confusions for listening in both quiet and speech-spectrum noise revealed significantly different error patterns that were related to each vocoder's ability to produce tone or noise output that accurately reflected the consonant's manner of articulation. Subject experience was also shown to influence intelligibility. Simulations using a computational model of modulation detection suggest that the noise vocoder's disadvantage is in part due to the intrinsic temporal fluctuations of its carriers, which can interfere with temporal fluctuations that convey speech recognition cues. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2773993]

I. INTRODUCTION

Researchers have long sought to determine which acoustic features of speech aid speech recognition in favorable and unfavorable listening conditions. One frequently investigated factor is the degree of spectral resolution required to produce intelligible speech. Experiments assessing the effects of spectral resolution often use speech synthesized by a channel vocoder, comprised of sums of amplitude-modulated carriers (e.g., pure tones or band-limited noise). Most of these experiments have considered the effects of only one carrier type; few have compared two or more carrier types.

This study compares the effects of pure-tone carriers and noise-band carriers on vocoded speech intelligibility. The results of this work are particularly relevant for cochlear implant processing. Similarities between channel vocoders and front-end processing for cochlear implants have led several investigators (Shannon *et al.*, 1995; Dorman *et al.*, 1997; Dorman and Loizou, 1998; Dorman *et al.*, 1998; Fu *et al.*, 1998; Loizou *et al.*, 1999; Friesen *et al.*, 2001; Qin and Oxenham, 2003) to use channel-vocoded speech as a model for cochlear implant processed speech. Researchers comparing implant and simulator performance have consistently shown that intelligible speech can be produced by vocoders with a small number of channels under favorable listening conditions. Shannon *et al.* (1995) showed that experienced subjects could correctly recognize at least 90% of medial vowels, medial consonants, and sentences produced by a four-channel noise-band vocoder. Similarly, the subjects of Dorman *et al.* (1997) achieved recognition scores of 90%

correct or better on similar listening materials processed by six-channel noise-band and pure-tone vocoders. Significant differences between the two vocoder types were found in only two cases: recognition of multi-talker vowels (favoring the sine-wave processor), and reception of place of articulation of consonants (favoring the noise-band processor).

The similarities in results of sine-wave and noise-band vocoder studies are somewhat unexpected, since the two types of vocoders produce signals with vastly different subjective characteristics. Given the practical challenges of synthesizing fricatives and bursts with only sine waves (George and Smith, 1997), one might expect to improve the performance of the vocoder by combining two carrier types as Dudley (1939) did in the first channel vocoder system. Toward this end, Whitmal *et al.* (2004) conducted a pilot experiment comparing the intelligibility of speech from six-band tone and noise-band vocoders with speech from a hybrid six-band vocoder using tone carriers in the lowest three bands and noise carriers in the highest three bands. This carrier allocation was intended to improve the quality of phonemes having primary spectral emphasis at either low frequencies (e.g., vowels) or high-frequencies (e.g., fricatives). Results indicated that sentences in quiet and in noise processed by the tone vocoder were more intelligible than those from the hybrid vocoder, which in turn were more intelligible than noise-band vocoded sentences. This surprising result (discussed later in the paper) suggests that the two carrier types may not be interchangeable for cochlear implant simulation experiments.

Two other recent studies have also reported performance advantages for sine-wave vocoders. In pilot testing for vowel recognition and gender identification tasks, Fu *et al.* (2004)

---

found that normal-hearing subjects listening through one-band and four-band noise-band vocoders performed poorly relative to cochlear-implant users. When they replaced their noise carriers with tone carriers, their subjects obtained higher levels of performance that better resembled the performance of cochlear implant users. Gonzalez and Oliver (2005) directly compared tone-vocoded and noise-vocoded sentences in speaker and gender identification tasks and also measured significantly better performance with the tone-vocoded sentences. The superior performance of tone vocoding in these experiments was attributed to both an apparent preservation of pitch periodicity and the greater sensitivity of the auditory system to amplitude modulation of tones.

Similar vocoder simulations (Dorman and Loizou, 1998; Fu *et al.*, 1998; Friesen *et al.*, 2001; Qin and Oxenham, 2003) have also been used to understand how speech recognition performance for cochlear implants is affected by additive noise. These studies suggest that recognition performance improves as the spectral resolution (reflected by the number of channels) and/or the signal-to-noise ratio (SNR) of the stimuli increases. Qin and Oxenham (2003), in particular, found that competing speech and amplitude-modulated speech-spectrum noise were less efficient maskers for unprocessed sentences than speech-spectrum noise, and more efficient maskers for vocoded sentences using four, eight, or twenty-four bands. The authors suggested that their vocoders were removing cues to pitch perception that made it more difficult for subjects to discriminate between target speech and masking speech.

It is difficult to reconcile the large differences between tone and noise-band vocoders described earlier (Whitmal *et al.*, 2004; Fu *et al.*, 2004; Gonzalez and Oliver, 2005) with the finding of Dorman *et al.* that there were few significant differences between the vocoders. Moreover, while some simulation studies have explored the effects of noise on sine-wave vocoded speech (Dorman *et al.* 1998) and noise-band vocoded speech (Fu *et al.*, 1998; Friesen *et al.*, 2001; Qin and Oxenham, 2003), none has performed a direct comparison of the intelligibility of the two vocoder types for speech in noise. Hence, it is not clear whether any similarities between intelligibility scores for tone and noise-band vocoder-processed speech would be observed in adverse conditions. The purpose of this study was to determine whether tone- and noise-band vocoded speech signals were equally intelligible in both quiet and noisy conditions. The present study consists of four experiments. Experiments 1 and 2 compared the intelligibility of sentences and medial consonants (respectively) for the two vocoders in both quiet and noisy conditions. Experiment 3 tested closed-set consonant recognition extensively in both quiet and noisy conditions for two purposes: analyzing error recognition patterns for each vocoder type, and observing learning effects. Experiment 4 compared the sine-wave vocoder with three different types of noise-band vocoders to determine the effects of carrier envelope fluctuations on consonant recognition.

II. EXPERIMENT 1: INTELLIGIBILITY OF TONE-VOCODED AND NOISE-VOCODED SENTENCES IN QUIET AND IN BACKGROUND NOISE

The purpose of Experiment 1 was to determine whether tone-vocoded and noise-vocoded sentences were equally intelligible in two types of background noise.

A. Methods

1. Subjects

Twelve adult female listeners participated in Experiment 1. The subjects' ages ranged from 21 to 38 years (mean age=25.8 years). All subjects were native speakers of American English with normal hearing. None of the subjects had participated in previous simulation experiments. The subjects were compensated for their participation with either a cash payment or partial course credit.

2. Materials

Stimuli for Experiment 1 consisted of 360 sentences, each containing three key words (Helfer and Freyman, 2004). The key words were one- or two-syllable nouns or verbs from the Francis and Kucera (1982) list of most common words. The sentences were assigned to one of 24 topics (e.g., food, clothing, politics) used to help listeners direct their attention to the target speaker when sentences were heard in the presence of competing speakers. The sentences were uttered by a female speaker with an American English dialect and digitally recorded in a sound-treated booth (IAC 1604) with 16 bit resolution at a 22 050 Hz sampling rate. The COOL EDIT PRO software package (Syntrillium Software, Phoenix, AZ) was used to scale each sentence to the same overall rms level (−24 dB relative to the maximum level for 16 bit resolution). Recognition data acquired in speech-spectrum noise for a previous study (Helfer and Freyman, 2004) were then used to group the 360 sentences into 24 15-sentence equally intelligible lists with average *unprocessed* key word recognition of 50% correct.

3. Processing

Subjects listened to vocoded versions of the above-described sentence materials, presented either as recorded in quiet or with masking noise added at SNRs of 3, 8, 13, or 18 dB before vocoding. Six-channel vocoders were chosen because they approach asymptotic performance in vocoder simulations without imposing ceiling effects (Dorman *et al.*, 1997), correspond to the effective number of channels that many cochlear-implant listeners can access (Dorman *et al.*, 1998), and represent a configuration for which simulation results are similar to results from cochlear-implant systems (Dorman and Loizou, 1998; Friesen *et al.*, 2001).

Two types of masking noise were used: speech-spectrum noise (SSN) and two-talker babble (TTB). The SSN was developed by creating a 110 250 sample white-noise signal (i.e., 5 s at a 22 050 Hz sampling rate), passing it through a 50th-order all-pole filter matched (via Levinson's recursion) to the average autocorrelation function for the 360 sentences, and scaling the filter's output to the average rms level of the 360 sentences. The TTB (used in Freyman *et al.*, 2001) con-

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Whitmal *et al.*: Vocoder carrier effects on speech intelligibility   2377

TABLE I. Band parameters for the six-channel vocoder.

Band	1	2	3	4	5	6
Center freq (Hz)	180	446	885	1609	2803	4773
Bandwidth (Hz)	201	331	546	901	1487	2453

sisted of digital recordings (sampling rate: 22 050 Hz) of two college-aged female students speaking different sets of nonsense sentences that were syntactically correct but semantically anomalous. Pauses between sentences were removed to produce two recordings of continuous speech which were then matched in rms level and combined to produce TTB. A 35-s-long stream of the babble was extracted for use in the present investigation.

Maskers for each sentence were derived from segments of either the SSN or TTB. Each segment was the same length as the individual sentence recordings, which each contained a short silent period of variable length both before (estimated mean=0.064 s) and after (estimated mean=0.328 s) the sentence was spoken. The origins of the masker segments were selected at random to prevent the subjects from hearing the same phrases or sections repeatedly. The masker segment was scaled to produce the desired SNR for the particular trial, and the target and masker were summed together before being input to the vocoder under test.

The vocoders used in the experiment were implemented via custom MATLAB software (Mathworks, Natick, MA) using the configuration of Qin and Oxenham (2003). The vocoders filtered their inputs (using sixth-order Butterworth filters) into six contiguous frequency bands in the 80–6000 Hz range. The frequency range was divided equally in terms of the Cam (or equivalent rectangular bandwidth) scale (Glasberg and Moore, 1990), such that each band was approximately 4.65 Cams in width. Bandwidth and band center frequencies are shown in Table I. It should be noted that the center frequency and bandwidth of Band 1, while lower than that used by Dorman et al. (1997), are consistent with frequency table parameters for the Nucleus-22 cochlear implant using SPEAK processing (Fu and Shannon, 1999).

Envelopes for each frequency band were obtained from filtered speech by half-wave rectification followed by smoothing with a second-order Butterworth low-pass filter. The bandwidth of the smoothing filter was the smaller of either 300 Hz or half the analysis bandwidth. The resulting envelopes were used to amplitude-modulate one of two carriers: a sine wave (located at the band's center frequency), or white noise subsequently filtered by that band's bandpass filter. The modulated carriers were level matched to their original in-band input signal and summed to produce simulated implant-processed speech.

4. Procedure

Subjects listened to the 24 lists of sentences while seated in a double-walled sound-treated booth (IAC 1604) during one 75 min listening session. Subjects were given breaks between the 12th and 13th lists. Prior to testing, each of the 24 lists was assigned to one of two groups; one for SSN, and one for TTB. Carrier/list pairings for the 12 subjects were

determined by two 12×12 Latin squares: one for the SSN group, the other for the TTB group. Each combination of the six SNRs and two carrier configurations was used in processing one of the twelve 15-sentence lists in a group. Sentences within each list were shuffled for presentation in random order. Each subject listened to all of the sentences. Half of the subjects listened to the SSN sentences before listening to TTB sentences; the remaining subjects listened to TTB sentences before listening to SSN sentences.

Custom MATLAB software (executed on a remote computer) was used to present the sentences to the subject and to score the number of key words correctly repeated by her. A laptop screen located inside the test booth prompted the subject with the word "Ready?" and the sentence topic exactly 2 s before the sentence was presented. The sentence was then retrieved from the remote computer's hard disk, converted to an analog signal by the computer's sound card (Analog Devices, SoundMax Integrated Digital Audio) using 16 bit resolution at a 22 050 Hz sampling rate, passed through an attenuator (Tucker-Davis PA4) and headphone amplifier (Tucker-Davis HB5), and presented diotically to the subject at 65 dBA through TDH-50P headphones. (TDH-series headphones have been used to present high-intelligibility vocoded speech in previous studies, e.g., Shannon et al., 1995; 1998.) Presentation levels were calibrated daily using repeated loops of the speech-spectrum noise described earlier. Upon hearing each sentence, the subject repeated what she heard into a talk-back microphone monitored by a researcher, who then recorded the number of key words correctly repeated.

Training materials were limited to ten sentences per carrier type presented without feedback just before starting the experiment. Of the ten sentences, five were presented in quiet and five were presented in speech-spectrum noise at 8 dB SNR. The sentences used for training were not used in the main experiment.

B. Results

Intelligibility scores for Experiment 1 were derived from the percentage of correctly repeated key words per condition. Mean sentence intelligibility scores for SSN and TTB are shown in Fig. 1. Under all test conditions, tone-vocoded sentences were more intelligible than noise-vocoded sentences. Scores for masking with speech-spectrum noise were higher than scores for masking with babble; this is consistent with the results of Qin and Oxenham (2003).

Subject scores were converted to rationalized arcsine units (Studebaker, 1985) and input to a repeated-measures analysis of variance (ANOVA) of intelligibility scores. Within-subject factors for the ANOVA included carrier configuration, noise type, SNR, and carrier presentation order (i.e., first group or second group). Results of the ANOVA indicated that carrier configuration ($F[1,287]=240.62$, $p < 0.0001$), noise type ($F[1,287]=21.43$, $p < 0.0001$), SNR ($F[5,287]=377.39$, $p < 0.0001$), and carrier order ($F[1,287]=11.50$, $p=0.0008$) were all significant main factors. Post hoc tests using Tukey's Honestly Significant Difference (HSD) criterion ($\alpha=0.05$) indicated that (a) tone-based voc-
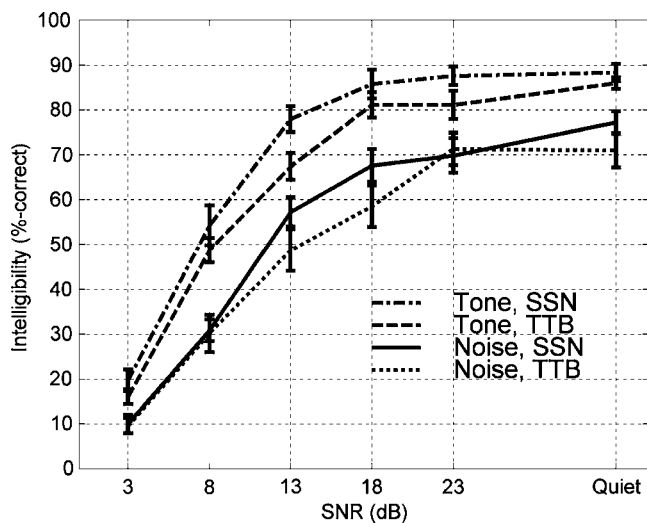
FIG. 1. Experiment 1: Percent-correct recognition scores for words in sentences (±1 SE) mixed with either speech-spectrum noise (SSN) or two-talker babble (TTB) when processed by tone and noise-band vocoders.

oding was significantly more intelligible than noise-band vocoding, (b) TTB was a significantly better masker than SSN, (c) scores at SNRs below 23 dB were significantly different from each other, and (d) average scores for the second carrier presented were significantly greater than average scores for the first carrier. A comparison of first-presented and second-presented means using Tukey's HSD criterion ($\alpha=0.05$) showed significant gains of 9.5 rational arcsine units (RAU) for both carrier types at 18 dB SNR. Smaller gains ranging between 3.1 and 5.7 RAU were also observed for both carriers at 8 and 13 dB SNR.

III. EXPERIMENT 2: INTELLIGIBILITY OF TONE-VOCODED AND NOISE-VOCODED CONSONANTS IN BACKGROUND NOISE

The tone-vocoded stimuli of Experiment 1 provided approximately 13% higher average intelligibility scores in quiet and approximately 20% higher average scores in masking noise than did the noise-vocoded stimuli. This contrasts markedly with previous research (Dorman *et al.*, 1997). The purpose of Experiment 2 was to determine whether the advantage for tone-vocoded stimuli extends to vowel-consonant-vowel (VCV) syllables, which are free from the syntactic, semantic, or contextual cues provided by the sentences of Experiment 1.

A. Methods

1. Subjects

Six adult female listeners participated in Experiment 2. The subjects' ages ranged from 21 to 38 years (mean age =25.5 years). Five of the subjects were completely inexperienced, having never participated in simulation experiments, and met the screening requirements of Experiment 1. The sixth subject was a participant in Experiment 1 who had no previous experience listening to processed consonant stimuli. The subjects were compensated for their participation with either a cash payment or partial course credit.

2. Materials

Stimuli for Experiment 2 consisted of the 23 consonants /b d g p t k f θ v ð h s ʃ z ʒ tʃ ʤ m n w l j r/, uttered in /a/C/a/ format. The consonants were spoken by a female talker with an American English dialect and digitally recorded in a sound-treated booth (IAC 1604) with 16 bit resolution at a 22 050 Hz sampling rate. As in Experiment 1, consonants were scaled to the same overall rms level.

3. Processing

Subjects listened to vocoded versions of the VCV materials described earlier, presented either as recorded or with the speech-spectrum noise of Experiment 1 added at SNRs of 3, 8, 13, 18, or 23 dB before vocoding. Listening in TTB (the more efficient masker) was eliminated in order to provide a wider range of scores. Vocoders included the tone- and noise-band vocoders of Experiment 1.

4. Procedure

Subjects listened to the VCV syllables while seated in a double-walled sound-treated booth (IAC 1604) during two 1-h listening sessions. Custom MATLAB software was used to present the stimuli to the subjects and to score the number of correctly recognized consonants. The laptop screen displayed a 5×6 grid of buttons, 23 of which contained the English spelling of one of the tokens, e.g., "asha." The unvoiced and voiced consonants /θ ð/ were represented by "atha" and "aTHa," respectively. The screen also contained an indicator sign which prompted the subjects by changing color and presenting the messages "Idle," "Ready?," and "Guess!"

Each trial consisted of two repetitions of each of the 23 consonants, presented in random order. Trials were presented in blocks of two, matched for SNR, with either tone-vocoding first and noise-vocoding second (TN), or noise first and tone second (NT). Subjects listened to one 12-block sequence in each session, with block types alternated (e.g., TN/NT/TN/NT, etc.) Half of the subjects heard sequences beginning with TN in session 1 and NT in session 2; the other half heard NT first in session 1 and TN first in session 2. This alternation method was intended to prevent the subject from having a persistent experience advantage on either vocoder. In order to acclimate subjects to the task, the first block in a sequence was always presented in quiet; SNRs for the remaining blocks were assigned at random without replacement.

The subject was prompted to start each trial with a mouse click when ready. Two seconds prior to each VCV's presentation, the indicator sign turned red and presented "Ready?" to the subject. The VCV was then sent to the computer's sound card (Silicon Integrated Systems 7018) using 16 bit resolution at a 22 050 Hz sampling rate, passed through a custom-built impedance matching network, and presented to the subject's right ear over TDH-50P headphones at 65 dBA. The indicator turned green and presented "Guess?" to the subject. The subject double-clicked on the button labeled with the VCV that she believed she heard. The indicator state returned to "Idle" for 2 s before returning to the "Ready?" state for the next consonant in the trial.
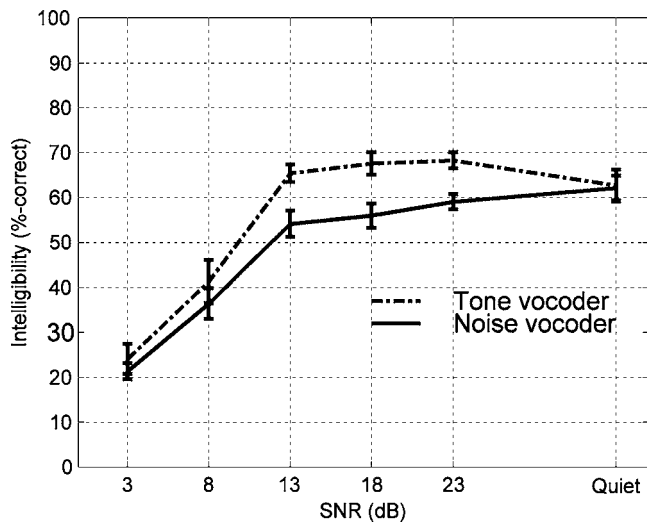
J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Whitmal *et al.*: Vocoder carrier effects on speech intelligibility   2379

FIG. 2. Experiment 2: Percent-correct recognition scores for vowel-consonant-vowel (VCV) syllables (±1 SE) mixed with speech-spectrum noise when processed by tone and noise-band vocoders. Two blocks of VCVs were allotted to each listening condition.

Training was conducted in two parts using VCVs uttered by a male speaker. First, one repetition of the unprocessed VCV set was presented in quiet. This was intended to (a) familiarize the subject with the response process and (b) make sure that the subject could identify and distinguish the consonants. The lowest score for any of the six subjects was 21 of 23 correct, or 91% correct. Typical errors for these subjects were confusions between different nonstrident fricatives with relatively low intensity and spectral peaks outside the composite bandwidth of the processor/headphone combination (Kent and Read, 2002). Next, four processed VCV sets were presented to the subject to provide limited practice for the main experiment. Each processed set used a unique combination of vocoder (tone or noise) and noise condition (in quiet or at 13 dB SNR).

B. Results

Intelligibility scores for Experiment 2 were derived from the percentage of correctly identified consonants per condition. A test-retest analysis of first- and second-session scores indicated excellent reliability, with significant intersession Pearson product-moment correlations for both vocoders (r =0.8281 for noise-vocoded speech and r=0.9254 for tone speech with $p<0.0001$). It is important to note that these high correlations only suggest a predictable linear relationship between first- and second-session scores; they do *not* imply that second-session scores are replications of first-session scores (measurement error notwithstanding). On average, Session 2 scores (52.4% correct) were slightly higher than Session 1 scores (50.6% correct), suggesting the possibility of a learning effect.

Mean consonant intelligibility scores for tone-based vocoders and noise-based vocoders are shown in Fig. 2. As in Experiment 1, tone-vocoded speech was more intelligible than noise-band-vocoded speech at SNRs of 8, 13, 18, and 23 dB. Scores for the two vocoders at 3 dB SNR were approximately equal at 23% correct; this likely reflects a floor

effect in the data. Scores in quiet were also approximately equal at 62% correct; this likely reflects a training effect (described in the following). Subject scores converted to RAU were input to a repeated-measures ANOVA with main factors of carrier configuration, SNR, and session number (i.e., first or second). Results indicated that carrier configuration ($F[1,143]=43.28$, $p<0.0001$) and SNR ($F[5,143] =187.28$, $p<0.0001$) were both significant main factors. The interaction between SNR and carrier (caused by the near-equality of scores in quiet and at 3 dB SNR) was also significant ($F[5,143]=3.38$, $p=0.0078$). No other interactions were significant. Session number approached (but did not reach) statistical significance ($F[1,143]=3.71$, $p=0.0573$). Post hoc tests using Tukey's HSD criterion ($\alpha=0.05$) indicated that (a) tone-based vocoding was again significantly more intelligible on average than noise-band vocoding, and (b) scores in quiet were significantly higher than those in noise at 3 or 8 dB SNR.

In Experiment 1, tone-based vocoding of quiet speech produced the most intelligible stimuli presented to the subjects. While the score of 62% correct in quiet for tone vocoding is a relatively high score in Experiment 2, it does not again represent the most intelligible of the eight conditions. One possible explanation for this interexperiment difference is that the higher intelligibility of the quiet speech was compromised by making the quiet speech the first signal presented in every session. This ordering would prevent the listener from applying any benefits of experience to the quiet speech. At the same time, the randomized presentation order used for speech in noise would allow listeners to apply varying amounts of experience to each trial. On average, then, vocoding in moderate SNRs like 18 dB might produce better performance than expected relative to scores for vocoding in quiet. At the same time, the combination of highly intelligible quiet speech in the first trial followed by less intelligible speech in successive trials might have confounded any existing learning effects. To explore these possibilities, a second ANOVA was conducted using only scores for speech in noise. Results indicated that removing quiet scores marginalized the SNR-carrier interaction ($F[4,119]=2.41$, $p =0.0573$) and made the advantage of Session 2 scores over Session 1 scores (now 50.9% correct to 48.4% correct) statistically significant ($F[1,119]=5.63$, $p=0.0205$).

IV. EXPERIMENT 3: INVESTIGATION OF ERROR PATTERNS AND LEARNING EFFECTS FOR TONE-VOCODED AND NOISE-VOCODED CONSONANTS IN QUIET AND IN NOISE

The data from Experiments 1 and 2 show an intelligibility advantage for tone-vocoded stimuli, both in quiet and in masking noise, and suggest the possibility of significant learning effects. The purpose of Experiment 3 was to explore the intelligibility advantage and learning effects in greater detail. Tone-vocoded and noise-vocoded VCVs at 8 dB SNR and in quiet were presented to the subjects of Experiment 2. The 8 dB SNR condition was chosen as a representative noise condition that was challenging for subjects but free from floor effects.
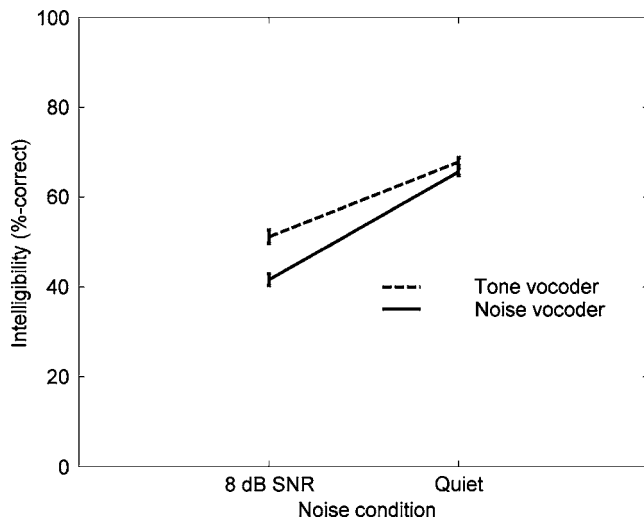
Whitmal *et al.*: Vocoder carrier effects on speech intelligibility

FIG. 3. Experiment 3: Percent-correct recognition scores (±1 SE) for Experiment 2 subjects listening to tone-vocoded and noise-vocoded VCV syllables, either in quiet or mixed with speech-spectrum noise at 8 dB SNR. Five blocks of VCVs were allotted to each listening condition.

A. Methods

1. Subjects

The six adult female listeners of Experiment 2 also participated in Experiment 3.

2. Materials

Subjects listened to tone- and noise-vocoded versions of Experiment 2 consonant recordings, presented either in quiet or in speech-spectrum noise at 8 dB SNR as in Experiments 1 and 2.

3. Procedures

Subjects listened to VCVs under the conditions of Experiment 2 in two listening sessions. The first session was dedicated to listening in speech-spectrum noise; subjects were later asked to return for listening in quiet. Each session consisted of two 10-block sequences, with block types alternated as in Experiment 2 (e.g., TN/NT/TN/NT, etc.). Subjects were given a break between sequences. This protocol allotted a total of five blocks (230 VCVs) to each condition (Experiment 2 allotted 92 VCVs), providing subjects with more rapid and extensive exposure to each condition than allowed in Experiment 2.

B. Results

1. Intelligibility scores

Intelligibility scores for Experiment 3 were derived as in Experiment 2. Mean consonant intelligibility scores for tone-based vocoders and noise-based vocoders are shown in Fig. 3. Two similarities between these data and those of Experiment 2 are noted. First, tone-vocoded speech was more intelligible than noise-vocoded speech. Differences between the carriers were most noticeable at 8 dB SNR, with average tone-vocoded scores of 51.2% correct and average noise-vocoded scores of 41.6% correct observed. Second, scores in

TABLE II. Linear regression model of subject learning effects, line parameters, correlation coefficients (r), and significance level (p).

Listening condition		Slope	Intercept	r	p
Tone	Quiet	0.70	64.07	0.77	0.009
vocoder	8 dB SNR	0.72	47.17	0.78	0.007
Noise	Quiet	0.56	62.57	0.64	0.045
vocoder	8 dB SNR	0.39	39.44	0.51	0.128

quiet were closer in value, with average tone-vocoded scores of 67.9% correct and average noise-vocoded scores of 65.6% correct observed.

Intelligibility scores for each vocoder/SNR combination were found to increase in near-linear fashion as a function of trial number as the subjects gained experience. Regression lines fit to data for each vocoder/SNR combination (see Table II) indicated that scores for tone-vocoded consonants increased at the approximate rate of 7 percentage points over ten trials. In contrast, scores for the more difficult noise-vocoded consonants increased at an approximate rate of 4.75 points over ten trials. The regression lines for the tone vocoder also fit the data better ($r=0.77$ in quiet and 0.78 in noise, $p<0.01$) than the lines for the noise vocoder ($r=0.64$ for quiet, $p=0.045$; $r=0.51$ for noise, $p=0.128$).

Subject scores were converted to RAU and input to a repeated-measures ANOVA with main factors of carrier configuration, SNR, and trial number. Results of the ANOVA indicated that carrier configuration ($F[1,239]=60.48$, $p<0.0001$), SNR ($F[1,239]=714.98$, $p<0.0001$), and trial number ($F[9,243]=3.17$, $p=0.0016$) were all significant main factors. The interaction between SNR and carrier (caused by the near-equality of scores in quiet) was also significant ($F[1,243]=21.38$, $p<0.0001$). Post hoc tests using Tukey's HSD criterion ($\alpha=0.05$) indicated that (a) tone vocoding was again significantly more intelligible than noise-band vocoding, (b) scores in quiet were significantly higher than those in noise at 8 dB SNR, and (c) trials 9 and 10 had significantly higher scores than trial 1 and trial 10 had significantly higher scores than trial 2.

2. Consonant confusions

Consonant confusion scores for Experiment 3 were derived from contingency tables (or, confusion matrices) comparing *a priori* voicing, manner, and place of articulation classifications with those produced by the subjects' identification tasks. Consonant confusions for voicing identification (shown in Table III) indicate nearly perfect performance in quiet: 98.7% correct for the tone vocoder versus 97.2% correct for the noise-band vocoder. In noise, errors increased in a complementary fashion, with the tone vocoder causing more unvoiced-to-voiced conversions and the noise vocoder causing more voiced-to-unvoiced conversions. The higher proportion of voiced phonemes (1680 vs 1080) ensured that tone vocoding maintained a small advantage over noise vocoding in noise (91% correct vs 86% correct).

Consonant confusions for manner identification (shown in Table IV) indicate good performance in quiet for both systems, with correct manner identification for 92% of noise-

TABLE III. Voicing confusion analysis of responses to VCV stimuli in Experiment 3.

	Tone vocoder		Noise vocoder	
Responses in quiet				
Stimulus	Unvoiced	Voiced	Unvoiced	Voiced
Unvoiced	1070	10	1058	22
Voiced	24	1656	53	1627
% correct	98.70%		97.20%	
Responses in noise (8 dB SNR)				
Stimulus	Unvoiced	Voiced	Unvoiced	Voiced
Unvoiced	895	185	958	122
Voiced	77	1603	259	1421
% correct	90.50%		86.20%	

vocoded VCVs and 94% of tone-vocoded VCVs. The largest difference between the systems was observed for correct classification of semivowels: 87% correct for noise vocoding, 94% correct for tone vocoding. In noise, performance dropped to 75% correct for noise vocoding and 79% correct for tone vocoding, with distinctive interaction between manner type and vocoder type. Tone carriers provided better recognition for stops, nasals, and semivowels; noise carriers provided better recognition for fricatives only. It is likely that the particular strengths of each vocoder were related to its carrier's ability to produce fricative-like noise. This phenomenon was also reflected in the error patterns for each vocoder.

TABLE IV. Manner confusion analysis of responses to VCV stimuli in Experiment 3.

Stimulus	Stop	Fricative	Affricate	Nasal	Semivowel
Tone vocoder responses in quiet					
Stop	719	1	0	0	0
Fricative	111	936	33	0	0
Affricate	2	3	235	0	0
Nasal	0	0	0	240	0
Semivowel	3	12	0	13	452
Score for this condition: 93.55% correct					
Noise vocoder responses in quiet					
Stop	716	2	2	0	0
Fricative	96	930	28	9	17
Affricate	0	5	234	1	0
Nasal	0	0	0	240	0
Semivowel	3	1	0	55	421
Score for this condition: 92.07% correct					
Tone vocoder responses in noise (8 dB SNR)					
Stop	642	39	13	24	2
Fricative	293	678	39	58	12
Affricate	1	3	236	0	0
Nasal	0	4	0	191	45
Semivowel	25	20	0	2	433
Score for this condition: 78.99% correct					
Noise vocoder responses in noise (8 dB SNR)					
Stop	463	160	90	2	5
Fricative	101	883	53	3	40
Affricate	3	3	234	0	0
Nasal	11	35	1	109	84
Semivowel	15	49	2	38	376
Score for this condition: 74.82% correct					

TABLE V. Place confusion analysis of responses to VCV stimuli in Experiment 3.

Stimulus	Labial	Dental	Alveolar	Palatal	Velar
Tone vocoder responses in quiet					
Labial	575	6	127	11	1
Dental	193	42	2	1	2
Alveolar	285	27	383	9	16
Palatal	50	0	53	495	2
Velar	25	1	8	0	446
Score for this condition: 70.3% correct					
Noise vocoder responses in quiet					
Labial	521	6	171	18	4
Dental	148	43	17	1	31
Alveolar	187	59	455	4	15
Palatal	34	0	44	521	1
Velar	67	3	33	0	377
Score for this condition: 69.5% correct					
Tone vocoder responses in noise (8 dB SNR)					
Labial	396	14	142	35	133
Dental	142	17	7	2	72
Alveolar	170	14	321	11	204
Palatal	30	1	65	486	18
Velar	8	0	7	3	462
Score for this condition: 60.9% correct					
Noise vocoder responses in noise (8 dB SNR)					
Labial	284	50	167	87	132
Dental	118	42	21	2	57
Alveolar	165	69	355	31	100
Palatal	35	8	44	481	32
Velar	75	25	80	19	281
Score for this condition: 52.3% correct					

The tone vocoder caused more fricative-to-stop conversions than the noise vocoder, which in turn converted more stops, nasals, and semivowels to fricatives than the tone vocoder. Similar results for noise vocoding were reported by Drullman *et al.* (1994), who suggested that the increased duration and modified envelopes of their medial stops led their subjects to sometimes identify them as fricatives.

Consonant confusions for place identification (shown in Table V) indicate poorer performance than for either manner or voicing identification in quiet or in noise. In quiet, tone vocoding and noise-band vocoding provided nearly equal accuracy (70.3% correct for tones, 69.5% correct for noise), with vastly different error patterns. Tone carriers provided better recognition for labial consonants (defined here as /p b m w f v/) and velar/glottal consonants, most of which were sonorants or stops with a low-frequency spectral emphasis. Conversely, noise carriers provided better recognition for alveolar consonants, most of which were fricatives or stops with a high-frequency spectral emphasis. Dentals (/θ ð/) and palatals (consisting largely of strident fricatives/affricates) were reproduced nearly as well by both vocoders. In noise, place identification dropped to 52.3% correct for noise vocoding and 60.9% correct for tone vocoding. Error patterns resembled those observed in manner identification: Tone vocoding made more dental-to-labial, dental-to-velar, and alveolar-to-velar conversions than noise vocoding, while noise vocoding made more velar-to-dental and velar-to-

TABLE VI. Kappa coefficients for feature confusion matrices of Experiment 3, with a χ^2 analysis testing the equality of coefficients for each feature.

Feature	Quiet		8 dB SNR		χ^2 for equal kappa test	p
	Noise vocoder	Tone vocoder	Noise vocoder	Tone vocoder		
Voicing	0.94	0.97	0.72	0.80	481.9	<0.001
Manner	0.89	0.91	0.66	0.72	590.4	<0.001
Place	0.60	0.61	0.39	0.50	240.2	<0.001

alveolar conversions than tone vocoding and converted 42% of labials to dentals, alveolars, or palatals. Of the remaining 39% of labials identified correctly, half were due to identification of or confusions between the fricatives /f/ and /v/. It should be noted that the nonstandard labeling of "labial" and "dental" phonemes used here allowed each phoneme to be identified by its voicing, manner, and place without creating more than five place designations.

The robustness of manner and voicing cues and relative fragility of place cues in both quiet and noise is expected, and consistent with results for both natural (Miller and Nicely, 1955) and vocoded (Dorman et al., 1997) speech. The exception to this trend is identification of the palatal consonants /ʃ ʒ tʃ dʒ r/, which were typically higher in level and/or longer in duration than either stops or nonstrident fricatives.

The chance-corrected agreement between stimuli and subject responses for each set of feature confusion matrices was quantified by computing the kappa coefficients (Cohen, 1960) shown in Table VI. Kappa coefficients are commonly used to describe the degree of agreement between two classification approaches, with values of kappa ranging between −1 (denoting no agreement) and +1 (denoting perfect agreement). Kappa values ranging from 0.8 to 1.0 denote near-perfect agreement between stimulus and response; from 0.6 to 0.8, substantial agreement; and from 0.4 to 0.6, moderate agreement (Landis and Koch, 1977). The computed kappa

FIG. 5. Experiment 4: Percent-correct recognition scores for vowel-consonant-vowel (VCV) syllables (±1 SE) mixed with speech-spectrum noise when processed by vocoders using tones, low-noise noise bands, wideband Gaussian noise, and narrow-band Gaussian noise as carriers.

values are consistent with the intelligibility data in Figs. 1–3, with better stimulus-response agreement shown for quiet conditions and tone vocoding. Performance for voicing and manner features, in particular, was excellent for both vocoders in quiet, while performance for place identification in quiet was only moderately accurate for both vocoders. Differences between noise and tone vocoding were most evident in noise, with good performance shown for voicing and manner and only moderate performance shown for place identification. Chi-squared tests (see Table VI) were conducted for each feature to test a null hypothesis of equal kappa values; the hypothesis was rejected in each case, further supporting the advantage of tone vocoding over noise vocoding.

V. EXPERIMENT 4: EFFECTS OF CARRIER BANDWIDTH AND ENVELOPE ON INTELLIGIBILITY OF VOCODED CONSONANTS

The data of Experiments 1, 2, and 3 indicate a performance advantage for tone-vocoded stimuli relative to noise-vocoded stimuli. Other researchers reporting performance advantages for tone vocoders (Fu et al., 2004; Gonzalez and Oliver, 2005) suggest that noise carrier envelope modulations reduce intelligibility by impairing detection of speech signal envelope modulations. This suggestion is supported by the computational model of Dau et al. (1999), which predicts higher modulation detection thresholds (MDTs) for noise carriers than for the tone carriers. The purpose of Experiment 4 was to investigate the relationship between the intelligibility of vocoded speech and predicted MDTs. The performance of the noise vocoder of Experiments 1 and 2 was compared to that of two additional noise-band vocoders using carriers examined by Dau et al. (1999). One of the two additional vocoders used narrow-band Gaussian noise carriers found to impair modulation detection; the other used "low-noise noise" carriers (Pumplin, 1985; Kohlrausch et al., 1997) found to facilitate modulation detection. Predictions of detection thresholds for the carriers were computed and analyzed in association with intelligibility scores.

A. Methods

1. Subjects

Twelve adult female listeners participated in Experiment 4. The subjects' ages ranged from 19 to 25 years. None of the subjects had participated in previous simulation experiments. The subjects were compensated for their participation with either a cash payment or partial course credit.
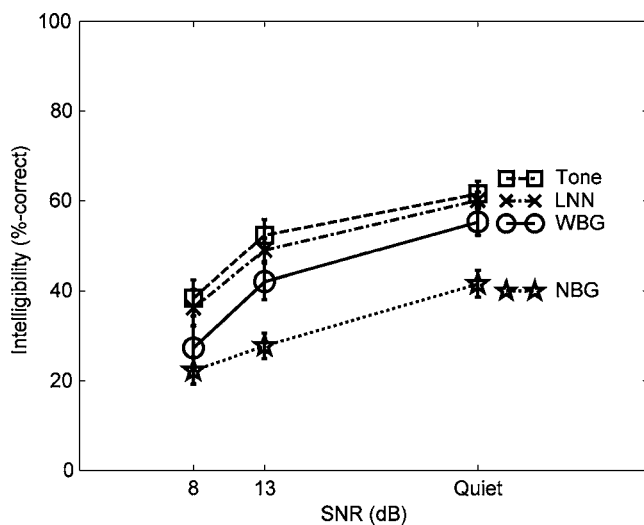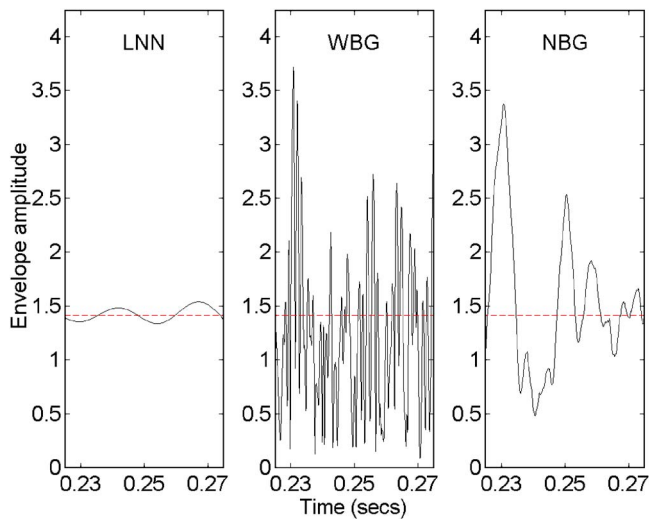
FIG. 4. (Color online) Experiment 4: 50 ms segments of low-noise noise, wideband Gaussian noise, and narrow-band Gaussian noise carrier envelopes in vocoder band 4 at equal rms levels. The envelope of a Band 4 tone carrier (at the same rms level) is also plotted for reference in each panel (dotted line).

2. Processing

Subjects listened to Experiment 2 consonant materials, presented in quiet or with masking noise added at SNRs of 8 or 13 dB before vocoding. These two SNRs were chosen on the basis of data from Experiments 1 and 2 suggesting that the two SNR values were challenging to subjects and less susceptible to either ceiling or floor effects than the other values used.

Four six-channel vocoder configurations were used in the experiment. Two of the vocoders were the tone and wideband Gaussian (WBG) noise vocoders used in the previous three experiments. Carriers for the third vocoder consisted of low-noise noise (LNN) signals (prepared as in Kohlrausch *et al.*, 1997) with bandwidths of 100 Hz and center frequencies matched to those of the vocoder's channels. The resulting carriers contained envelopes that were nearly flat (like the tone carriers' envelopes) with narrow-band spectra. Carriers for the fourth vocoder were narrow-band Gaussian (NBG) noise signals, created by passing white noise through a bank of fourth-order Butterworth filters with bandwidths of 100 Hz and center frequencies matched to those of the vocoder's channels. Representative noise carrier envelope wave forms are shown in Fig. 4. As expected, the LNN carrier presented envelopes with minimal temporal fluctuation, while the Gaussian carriers presented rapidly fluctuating envelopes with spectral energy at frequencies above the speech modulation frequency range (Houtgast and Steeneken, 1985).

3. Procedure

Subjects listened to the VCVs under the conditions described for Experiment 2, with two variations. First, the training session was used to identify viable subjects. As in Experiment 2, one repetition of the unprocessed consonant set was presented in quiet to each subject. Nine of the twelve subjects achieved scores of 87% correct or better. Error pat-

terns for these subjects resembled those of subjects in Experiment 2. The remaining three subjects achieved low scores (65%, 78%, and 78% correct) that reflected poor consonant recognition, and were excused from further participation. Four processed consonant sets in quiet (one per vocoder type) were then presented to the remaining subjects for limited practice. Second, trial ordering for each subject followed an extended form of the ordering scheme of Experiment 2 in order to reproduce (as much as possible) that experiment's learning effects. The twelve subjects were first counted off into four groups of 3 to determine presentation order. Trials were presented in blocks of four (one block per vocoder, as in Experiment 2), matched for SNR. The ordering of the vocoders in each block was based on a fixed sequence of four numbers per subject, with the first number equal to the subject's group number, and the other three following in random order. (Example: A viable sequence for a subject in Group 1 might be {1 4 2 3}). Subjects listened to three 4-block sequences in each session, with the first and third sequences following their set order and the second sequence in reverse order. (Example: Using the above-noted sequence, Subject 5 would hear vocoders in the following order: 1 4 2 3 3 2 4 1 1 4 2 3.) The first block in a sequence was always presented in quiet; SNRs for the remaining blocks were assigned at random without replacement.

B. Results

1. Intelligibility scores

Consonant intelligibility scores for Experiment 4 were derived as in Experiment 2. Mean consonant intelligibility scores for the four vocoders are shown in Fig. 5. As before, tone vocoded speech was more intelligible than WBG noise-vocoded speech. Average scores for the LNN vocoder, however, were only 2.1 percentage points below those of tone-vocoded speech. Speech from the NBG vocoder was least intelligible of all, with its average scores in *quiet* (42% correct) measuring well below average scores at 13 dB SNR (58% correct) for the other three vocoders.

Subject scores were converted to RAU and input to a repeated-measures ANOVA with main factors of carrier configuration and SNR. Results indicated that carrier configuration ($F[3,48]=52.39$, $p<0.0001$) and SNR ($F[2,48]=119.03$, $p<0.0001$) were both significant main factors. No significant interactions were observed. Post hoc tests using Tukey's HSD criterion ($\alpha=0.05$) indicated that (a) results for tone-based vocoding and LNN-based vocoding were not significantly different, (b) speech from the tone and LNN vocoders was significantly more intelligible than the two Gaussian noise vocoders, (c) speech from the wide-band Gaussian noise vocoder was significantly more intelligible than the narrow-band Gaussian noise vocoder, and (d) average scores at all SNRs were significantly different from each other.

Figure 5 also reveals that intelligibility scores for tone and WBG noise vocoders in Experiment 4 were noticeably lower than comparable scores in Experiment 2. The largest discrepancies between the two experiments were observed at 13 dB, the SNR value least susceptible to either floor or

ceiling effects. An inspection of individual scores revealed that intelligibility for most subjects in Experiment 2 tended to increase monotonically with SNR, while several Experiment 4 subjects did not show monotonic increases and often performed far worse than their peers at either 8 or 13 dB. One subject in particular (JW4) produced an average score of 28.6% correct, with lower scores in nine conditions than the other eight Experiment 4 subjects (who averaged 42.8% correct) and lower scores (average: 26.8% correct) in all comparable conditions than the six Experiment 2 subjects (average: 53.68% correct) who also listened to tone and wideband Gaussian noise vocoders in quiet and at 8 and 13 dB SNR. Subject scores from Experiments 2 and 4 were subsequently input to a pair of pooled ANOVAs with main factors of carrier, SNR, and experiment number (i.e., 2 or 4), computed both with and without scores from JW4. Significant interexperiment average differences were seen in both cases: 6.3 RAU when JW4's scores were included, and 3.8 RAU when excluded. In the latter case, Tukey's HSD criterion showed no significant differences between the average scores of individual subjects. Moreover, the largest interexperiment differences were observed with noise-band vocoders; these may be attributed in part to subject experience. In Experiment 2, subject experience was divided between two different vocoders (tone and WBG). In Experiment 4, subject experience was divided among four different vocoders (tone, LNN, WBG, and NBG), with the two most intelligible vocoders sounding (and performing) similarly. The subjects' remaining attention was split between the WBG and the less intelligible NBG vocoder. Reducing the time that the subject spent listening to the WBG vocoder may (as suggested by Experiment 3) have prevented subjects from acclimating to the WBG vocoder over time as well as they may have acclimated to the tone vocoder.

2. Predicted modulation detection thresholds

MDT predictions were computed using the algorithm of Dau et al. (1999). Dau et al. (1997a, b) had modeled modulation detection in normal-hearing subjects with a signal processor consisting of four stages: peripheral auditory filtering, envelope detection, nonlinear amplitude compression, and a modulation-frequency bandpass filter bank. The processor's responses to probe signals were used as templates in a signal detection stage that simulated subject responses in modulation detection and modulation masking experiments. The authors postulated that MDTs would be influenced by carrier envelope fluctuations transmitted to the output of each modulation-frequency filter. Their resulting simulations (Dau et al., 1997b) were in good agreement with their subjects' responses. In later work (Dau et al., 1999), the authors retained only the envelope detector and modulation filter bank, and modeled MDTs as the ratio of the average ac output power in the filtered envelope to the average carrier power. As before, they found good agreement between their predicted and measured data.

In the present work, carrier levels for the predictions were determined by passing an array of 50 sentences (scaled to 65 dB SPL) through each of the four vocoders and measuring the rms levels prescribed for the carriers in each of the
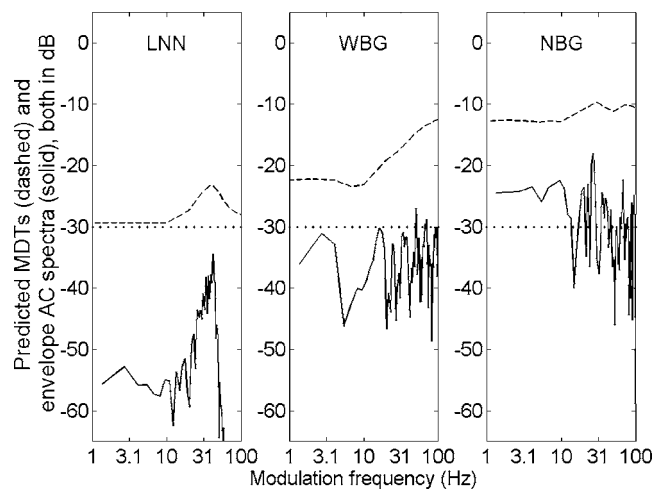

FIG. 6. Experiment 4: Predicted MDTs (upper dashed line) for low-noise noise (LNN), wideband Gaussian noise (WBG), and narrow-band Gaussian (NBG) noise carriers, plotted against modulation frequency. All carriers are limited to the range of vocoder band 4, a band exhibiting moderate differences between WBG and NBG thresholds. Thresholds are expressed as the ratio of the power in a sinusoidal modulation signal to the power of the carrier being modulated. The spectra of each carrier's unfiltered envelope fluctuations about the average envelope value are represented by curves with solid lines. For display purposes only, the envelope energy was normalized to a value of one to facilitate comparison with the MDTs. The MDT of a tone carrier (Dau et al., 1999) is represented by a dotted line.

six bands. Five-hundred millisecond segments of the 24 carriers were then generated, scaled to their appropriate rms level, and processed in three stages. First, carrier envelopes were computed from each carrier's Hilbert transform. Each envelope function was then input to a first-order modulation-frequency bandpass filter with complex-valued coefficients (Dau et al., 1997a) centered at one of several modulation frequencies (1, 2, 4, 5, 10, 20, 50, 70, or 100 Hz). The filtered-envelope-to-carrier power ratio was then computed as an estimate of the carrier's MDT. As in Dau et al. (1999), a constant value of 0.001 was added to the ratio to impose a threshold floor of −30 dB measured for sine carriers at moderate levels (Dau et al., 1997a).

Representative predicted MDT values for each carrier in Band 4 are indicated in Fig. 6 by dashed lines. The LNN carrier's MDTs are nearly identical to the tone carrier's −30 dB value at speech modulation frequencies (indicated by a dotted line), with a substantial threshold shift only seen near 40 Hz. In contrast, both the WBG and NBG carriers produce higher predicted MDTs than the LNN carrier. Differences between the MDT predictions are due to corresponding differences in the spectra of ac coupled envelope wave forms. Examples of wave form spectra for each carrier are indicated in Fig. 6 by solid lines. Since all carriers are normalized to the same rms level, the narrower bandwidth of the NBG carrier serves to concentrate more carrier envelope power at low modulation frequencies, which in turn lowers the envelope-to-carrier power ratio. As a result, the WBG carrier produces lower predicted MDTs than the NBG carrier, with differences ranging between 1.8 dB (band 1) and 9.0 dB (band 6). This effect is particularly evident in bands 1 and 2, where the 100 Hz NBG carrier bandwidth is compa-

rable to the width of the vocoder band and differences between WBG and NBG thresholds are smallest.

The data of Figs. 5 and 6 suggest a negative association between intelligibility scores and predicted MDTs. Computed Pearson product-moment correlation coefficients for intelligibility scores in quiet and 4 Hz detection threshold levels in individual bands 4, 5, and 6 support this observation ($-1.00 < r < -0.97$, $p < 0.025$). Note that the 4 Hz modulation frequency was selected as a representative frequency for speech modulations (Houtgast and Steeneken, 1985).

VI. DISCUSSION

A. Differences between tone and noise carriers

The results of the present study suggest that tone-vocoded speech and noise-vocoded speech are not equally intelligible in either quiet or noisy listening conditions for subjects with limited training. In Experiment 1, subject scores for a sentence identification task were significantly higher for tone-vocoded speech than for noise-vocoded speech, with measured differences of 13 percentage points in quiet and 17 percentage points in masking noise. Likewise, in Experiment 2, subject scores for a consonant identification task in masking noise were 8 percentage points higher (a significant difference) for tone-vocoded speech than for noise-vocoded speech. The results of Experiment 4 suggest that these differences between vocoders may be due in part to extraneous envelope fluctuations. Tone carriers have no intrinsic envelope fluctuations, and facilitate better modulation detection than noise-band carriers (Dau et al., 1999). As a result, tone carriers appear to be better at faithfully reproducing speech envelope fluctuations than noise-band carriers. This theory is supported by other studies (Cazals et al., 1994; Fu, 2002) showing negative associations between intelligibility scores and MDTs in cochlear implant users. It may also explain why the hybrid vocoder of Whitmal et al. (2004) with noise-band carriers in high-frequency bands did not perform as well as the tone vocoder. Conversely, it is plausible that noise-band carrier fluctuations were occasionally mistaken for actual speech envelope fluctuations. This phenomenon may explain why the Experiment 3 vocoders produced manner-of-articulation errors consistent with the ability to produce fricative-like noise. Specifically, the noise vocoder converted 35% of stop consonants into fricatives or affricates, while the tone vocoder converted 27% of fricatives into stops.

Masker envelopes may provide another source of extraneous envelope fluctuations. Results from Experiments 1 and 2 showed that TTB is a more efficient masker of vocoded speech than speech-spectrum noise. Other researchers obtaining similar results with speech maskers (Kwon and Turner, 2001; Qin and Oxenham, 2003; Stickney et al., 2004; Nelson and Jin, 2004; Gonzalez and Oliver, 2005) suggest that the limited spectral resolution of vocoded speech forces listeners to rely more heavily on temporal cues that become degraded by the intrinsic envelope fluctuations of the noise-band carriers. In particular, Qin and Oxenham (2003) note that the vocoders' poor spectral resolution at low frequencies

hinders subjects from using target/masker pitch differences and/or comodulation masking release to separate the target from the masker.

Carrier sidebands are a second major source of difference between the tone and noise-band vocoders. Amplitude-modulated tone carriers have sidebands that (depending on auditory filter responses at carrier frequencies) can provide additional detection cues. The 300 Hz bandwidth of the present study's envelope smoothing filters is broad enough to pass pitch-related periodic temporal fluctuations into the modulation envelope. As a result, envelope spectrum components at the talker's average fundamental frequency (approximately 215 Hz) appear as carrier sidebands that are outside the passbands of auditory filters centered at the carrier frequencies for bands 1–5. These sidebands impose a periodic temporal structure on the vocoder's output, with the talker's pitch accurately replicated over a majority of voiced segments (de Cheveigné and Kawahara, 2002). Gonzalez and Oliver (2005), who used envelope smoothing filters with a minimum bandwidth of 160 Hz, observed similar replications of pitch structure in their tone-vocoded signals. In contrast, sidebands of noise-band carriers are masked by carrier spectra, such that no periodic temporal structure is produced.

B. Comparisons with previous vocoder implementations

The results of Experiments 1 and 2 contrast with the results of Dorman et al. (1997), who measured intelligibility scores above 90% correct and perfect manner transmission for their six-band tone and noise vocoders. In addition, place transmission was significantly better for their noise vocoder (76% correct) than for their tone vocoder (70% correct). They attributed the noise vocoder's advantage to its contiguous bands, which they argued were better suited than tones for transmitting some place cues. The differences between the data of Dorman et al. (1997) and the data in this study may be attributed in part to differences in vocoder spectral resolution and emphasis. Previous work (Dorman et al., 1998; Loizou et al., 1999) suggests that higher spectral resolution in the 900–2500 Hz region can improve perception of vocoded consonants and second formants. Shannon et al. (1998) reduced the intelligibility of speech for a four-band vocoder (Shannon et al., 1995) from 95% correct to 80% correct by increasing the tonotopic width of the band containing 1500 Hz (a selected boundary point) from 9.7% of basilar membrane length (Greenwood, 1990) to 12.8% and 14.7%. Likewise, the six-band Dorman et al. (1997) vocoder used tonotopic bandwidths of 8.6–9.4% in the 900–2500 Hz region, whereas the vocoder in the present work used bandwidths ranging between 10.6% and 11%. The front ends of the Shannon et al. (1995) and Dorman et al. (1997) vocoders also featured 1200 Hz high-pass pre-emphasis filters for limiting upward spread of masking. Their combinations of pre-emphasis and spectral resolution provide high-frequency emphasis that (in informal listening) produces a clearer, more intelligible signal that may account for the best scores of the vocoders of both this study and of Qin and Oxenham (2003) being significantly lower than the 90%-correct scores reported by Dorman et al. (1997). (It should be noted that Qin

and Oxenham did not publish their raw data; performance here is inferred from their published two-parameter cumulative Gaussian functions to which they fit individual subject data
(mean=SRT, SD=slope) for listening in either SSN or to a single-talker masker.)

C. Effects of subject training

The extent of the subjects' training may also have influenced the results of the present experiments. Other studies using more extensive training regimens reported higher intelligibility scores. Shannon *et al.* (1995) reported that their subjects practiced listening to the stimuli for 8–10 h, with reported scores above 90% correct reflecting stabilized performance. Dorman *et al.* (1997) provided subjects with two passes through all items with visual feedback, ran a sample test with visual feedback given after each answer, and then presented test conditions in increasing order of difficulty (i.e., nine channels first, followed by eight channels, then seven, etc.) in order to better familiarize subjects with the tasks. Their subjects also achieved sentence recognition scores greater than 90% correct. Loizou *et al.* (1999) later measured sentence intelligibility of only 63% correct with the same vocoder using target sentences uttered by 135 different talkers. The authors subsequently argued that the single-talker targets used in previous studies helped elevate intelligibility scores by eliminating the need for subjects to learn to adapt to varying stimuli. Stone and Moore (2003) presented subjects with five counterbalanced blocks of 45 sentences each in a sentence intelligibility task and observed a 41%-correct increase in average intelligibility between the first and fifth block presentations. Since most of the increase occurred within the first two blocks, they suggested that investigators provide between 30 and 60 min of training for subjects in future studies. Given the effects of training and experience, it is possible that a comparison of sine-wave and noise-band processed stimuli conducted without extensive training may be more sensitive to differences in intelligibility than that of Dorman *et al.*, particularly in noise where useful recognition cues are further obscured.

D. Implications for cochlear implant simulation

Finally, the choice of carrier type for optimal simulation of cochlear implant performance requires some discussion. The speech recognition capabilities of subjects in cochlear implant simulation studies using both tone carriers (Dorman *et al.*, 1998) and noise-band carriers (Fu *et al.*, 1998; Friesen *et al.*, 2001) have resembled those of the best cochlear implant users in these studies' patient populations. However, many contemporary cochlear implant processors excite electrodes with amplitude-modulated pulse train carriers, rather than tone or noise carriers. As a result, there are often substantial differences between auditory nerve responses to implant stimuli and acoustic stimuli from a vocoder (Litvak *et al.*, 2001) that prevent either scheme from precisely modeling implant performance.

One rationale for using tone vocoders as cochlear implant simulators comes from Dorman *et al.* (1997), who re-

ported that cochlear implant users perceived individual channel stimulation as tone-like percepts, rather than noise-like percepts. The gender/speaker identification studies of Fu *et al.* (1998) and Gonzalez and Oliver (2005) further suggest that the perception of tone-vocoded speech is better and more similar to that of cochlear implant users than noise-band vocoded speech. At the same time, the fine spectral resolution that makes speech from tone vocoders more intelligible prevents the vocoders from accurately modeling channel interaction effects. This is illustrated by the work of Fu and Nogaki (2005), who showed that noise-band vocoders with broadly overlapping channel filters were better matched to cochlear implant performance than vocoders with steeply sloping channel filters. The authors noted similarities between four-channel vocoders with steeply sloped filters and eight-channel vocoders with overlapping filters, suggesting that there may be a range of acceptable channel bandwidth/channel overlap simulator combinations, none of which can be realized with a tone vocoder. Further research will be required to develop simulation methods that properly incorporate the salient features of both vocoder types.

The results of the present study are consistent with those of Fu *et al.* (2004) and Gonzalez and Oliver (2005) favoring the fidelity of tone vocoders. However, this fidelity difference does not address the question of whether one vocoder is a better simulator of implant performance. Moreover, the small intersubject performance differences observed for either of the present vocoders do not accurately model the large intersubject performance differences observed within or across studies of actual implant users (Friesen *et al.*, 2001; Fu *et al.*, 1998; Stickney *et al.*, 2004). Differences observed across studies may be in part attributable to implant/vocoder parameter differences (e.g., frequency range, number/width of bands, ceiling/floor effects, frequency allocation tables). Given these differences, it is not necessarily possible to assess conditions within one study that closely match the processing characteristics of all active CI recipients. There is, however, much to be learned from vocoder experiments, particularly in light of the fact that researchers may recruit large numbers of subjects, evaluate parameters not easily manipulated in actual cochlear implant systems, and then select the most important and sensitive parameters for use in studies with actual cochlear implant recipients.

VII. SUMMARY AND CONCLUSION

Channel vocoders employing either tone or bandpass noise carriers are often used to simulate cochlear implant processing in normal-hearing listeners. Previous research has suggested that the two types of carriers provide similarly high levels of performance in vocoders with as few as four bands. The present work compared tone and noise vocoders with six bands in both quiet and noisy listening conditions with subjects who have not undergone extensive training. In all four experiments, vocoders using tone carriers produced more intelligible speech than vocoders using noise carriers. An analysis of consonant confusions indicated that recognition error patterns for the two types of vocoders were also significantly different. These differences in performance

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Whitmal *et al.*: Vocoder carrier effects on speech intelligibility   2387

were attributed in part to the noise carriers' intrinsic fluctuations, which can impair detection of envelope fluctuations produced by speech, and in part to sidebands imparting a periodic temporal structure in voiced speech segments. Differences between the present results and those of previous studies (Shannon *et al.*, 1995; Dorman *et al.*, 1997) were attributed in part to differences in vocoder parameters and subject training protocols. These factors typically vary widely from experiment to experiment. Future research directed at understanding the effects of these factors may result in improved models of cochlear implant processing and perception.

ACKNOWLEDGMENTS

Cazals, Y., Pelizzone, M., Saudan, O., and Boex, C. (**1994**). "Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants," J. Acoust. Soc. Am. **96**, 2048–2054.

Cohen, J. (**1960**). "A coefficient of agreement for nominal scales," Educ. Psychol. Meas. **20**, 37–46.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2982–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," J. Acoust. Soc. Am. **102**, 2906–2919.

Dau, T., Verhey, J., and Kohlrausch, A. (**1999**). "Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers," J. Acoust. Soc. Am. **106**, 2752–2760.

de Cheveigné, A., and Kawahara, H. (**2002**). "YIN, a fundamental frequency estimator for speech and music," J. Acoust. Soc. Am. **111**, 1917–1930.

Dorman, M. F., and Loizou, P. C. (**1998**). "The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels," Ear Hear. **19**, 162–166.

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.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am. **102**, 2403–2411.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053–1064.

Dudley, H. (**1939**). "Remaking speech," J. Acoust. Soc. Am. **11**, 169–177.

Francis, W. N., and Kucera, H. (**1982**). *Frequency Analysis of English Usage: Lexicon and Grammar* (Houghton Mifflin, Boston).

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

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

Fu, Q.-J., Chinchilla, S., and Galvin, J. J. (**2004**). "The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users," J. Assoc. Res. Otolaryngol. **5**, 253–260.

Fu, Q.-J., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.

Fu, Q.-J., and Shannon, R. V. (**1999**). "Effects of electrode configuration and frequency allocation on vowel recognition with the Nucleus-22 cochlear implant," Ear Hear. **20**, 321–331.

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.

George, E. B., and Smith, M. J. T. (**1997**). "Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model," IEEE Trans. Speech Audio Process. **5**, 389–406.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Gonzalez, J., and Oliver, J. C. (**2005**). "Gender and speaker identification as a function of the number of channels in spectrally reduced speech," J. Acoust. Soc. Am. **118**, 461–470.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Helfer, K. S., and Freyman, R. L. (**2004**). "Development of a topic-related sentence corpus for speech perception research," J. Acoust. Soc. Am. **115**, 2601–2602.

Houtgast, T., and Steeneken, H. J. M. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069–1077.

Kent, R. D., and Read, C. (**2002**). *The Acoustic Analysis of Speech* (Delmar/Thomson Learning, Albany, NY).

Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A., and Puschel, D. (**1997**). "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," Acust. Acta Acust. **83**, 659–669.

Kwon, B. J., and Turner, C. W. (**2001**). "Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference?," J. Acoust. Soc. Am. **110**, 1130–1140.

Landis, J. R., and Koch, G. G. (**1977**). "The measurement of observer agreement for categorical data," Biometrics **33**, 159–174.

Litvak, L., Delgutte, B., and Eddington, D. (**2001**). "Auditory nerve fiber responses to electric stimulation: Modulated and unmodulated pulse trains," J. Acoust. Soc. Am. **110**, 368–379.

Loizou, P. C., Dorman, M., and Tu, Z. (**1999**). "On the number of channels needed to understand speech," J. Acoust. Soc. Am. **106**, 2097–2103.

Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.

Nelson, P. B., and Jin, S.-H. (**2004**). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **115**, 2286–2294.

Nilsson, M., Soli, S., and Sullivan, J. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.

Pumplin, J. (**1985**). "Low-noise noise," J. Acoust. Soc. Am. **78**, 100–104.

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Shannon, R. V., Zeng, F-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Shannon, R. V., Zeng, F-G., and Wygonski, J. (**1998**). "Speech recognition with altered spectral distribution of envelope cues," J. Acoust. Soc. Am. **104**, 2167–2476.

Stickney, G. S., Zeng, F-G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.

Stone, M. A., and Moore, B. C. J. (**2003**). "Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task," J. Acoust. Soc. Am. **114**, 1023–1034.

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Whitmal, N. A., Poissant, S., Freyman, R. L., and Helfer, K. S. (**2004**). "Effect of combining different carriers across bands on speech intelligibility in cochlear implant simulation," 27th Annual Mid-Winter Meeting of the Association of Research in Otolaryngology, Daytona Beach, FL.

Static features in real-time recognition of isolated vowels at high pitch

Aníbal J. S. Ferreira[a]

Department of Electrical and Computer Engineering, University of Porto, Rua Dr. Roberto Frias s/n,
4200-465 Porto, Portugal

This paper addresses the problem of automatic identification of vowels uttered in isolation by female and child speakers. In this case, the magnitude spectrum of voiced vowels is sparsely sampled since only frequencies at integer multiples of F0 are significant. This impacts negatively on the performance of vowel identification techniques that either ignore pitch or rely on global shape models. A new pitch-dependent approach to vowel identification is proposed that emerges from the concept of timbre and that defines perceptual spectral clusters (PSC) of harmonic partials. A representative set of static PSC-related features are estimated and their performance is evaluated in automatic classification tests using the Mahalanobis distance. Linear prediction features and Mel-frequency cepstral coefficients (MFCC) coefficients are used as a reference and a database of five (Portuguese) natural vowel sounds uttered by 44 speakers (including 27 child speakers) is used for training and testing the Gaussian models. Results indicate that perceptual spectral cluster (PSC) features perform better than plain linear prediction features, but perform slightly worse than MFCC features. However, PSC features have the potential to take full advantage of the pitch structure of voiced vowels, namely in the analysis of concurrent voices, or by using pitch as a normalization parameter. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2772228]

I. INTRODUCTION

Language learning and speech therapy applications require robust automatic identification of speech sounds uttered in isolation, notably vowels. Most frequently, the automatic identification of isolated vowels is implemented through formant analysis using, for example, linear prediction (LP) techniques (Zahorian and Jagharghi, 1993). The basic assumption when using LP techniques is that the production of voiced sounds by the human phonetic system can be modeled as an all-pole filter that is excited by a periodic train of glottal pulses. The repetition rate of these pulses corresponds to the fundamental frequency (F0) of the voiced sound, or pitch, and the poles of the all-pole filter correspond to resonances of the vocal tract, or formants. The frequencies of the first three formants (F1, F2, F3) are usually considered as good acoustic correlates of a given vowel (Fant, 1970). When F0 is significantly lower than the lowest formant (F1), the formant frequencies give rise to clear peaks in the envelope of the magnitude spectrum of the vowel signal. LP analysis techniques can therefore be used to reliably estimate the resonant frequencies of the vocal tract (or formant frequencies). Subsequently, a simple acoustic-phoneme association can be used to identify the vowel (Rabiner and Juang, 1993). On the other hand, when F0 is comparable to or higher than F1, which is typical in female or child speech, or singing, LP techniques are not reliable in estimating the formant frequencies and in subsequently identifying vowels. In fact, the pitch frequency of female or child speech can be on the order of 300 Hz (or even more), which significantly ex-

ceeds the upper limit of the typical male pitch range (about 160 Hz), but is comparable to a typical frequency of the first formant (F1). In this case, the LP approach normally fails to provide a reliable estimate of the formant frequencies, essentially because the magnitude spectrum becomes undersampled (i.e., it is sampled at integer multiples of the pitch frequency). A frequent observation is that using LP, the estimated formant frequencies are "locked" to the harmonic frequencies of the fundamental frequency. In other words, as noted by Zahorian and Jagharghi (1993), Bladon (1982), and Mollis (2005), conceptually, the LP model is valid for speech production, but is not appropriate for speech perception, notably in the case of vowel identification at high pitch frequencies. This problem has been addressed by de Cheveigné and Kawahara (1999) as a problem of *missing-data model of vowel identification*. To a large extent, it also corroborates that the question formulated by Klatt (1982) of how the "central nervous system can examine an input vowel spectrum and select out the relevant acoustic details in order to make a phonetic judgement," remains significantly unanswered.

In this paper a new approach to the problem of automatic identification of voiced vowels is proposed, where pitch plays a central role. The underlying concept of vowel perception assumes that the human auditory system (HAS) discriminates vowels by performing a perceptual clustering of partials in the harmonic structure of the vowel. This concept is designated in this paper as perceptual spectral cluster (PSC). While the formant concept derives from speech production and is specifically associated with the resonant frequencies of the vocal tract, the PSC concept is linked to perception and attempts to identify clusters of harmonic par-

a)Electronic mail: ajf@fe.up.pt

tials whose features, namely center of gravity (COG), left and right borders, and average spectral power, give rise to relevant perceptual cues that are used by the HAS in vowel discrimination and identification. The focus is placed on static features since previous results, such as those presented by Zahorian and Jagharghi (1993), have indicated that static features are the most important features in vowel discrimination, and also because the main application objectives considered in this paper involve real-time visual representation of static features extracted from sustained vowel sounds, and real-time vowel identification. Therefore, it is presumed in this context that a voiced vowel is uttered in isolation and in a sustained way (i.e., there is silence before and after the vowel sound). The envisioned application scenarios imply that the accumulated delay due to the signal processing involved in signal analysis, feature extraction, vowel classification, and visual representation is commensurate with the time the HAS takes in identifying isolated vowels. For most sustained vowels, robust human recognition occurs very reliably in less than 100 ms.

A. Motivation to real-time isolated vowel recognition

An interest in real-time robust vowel identification emerged when the author was involved in a project aiming at building a set of interactive demonstrators on the subject of sound and the human auditory system.[1] One such specific demonstrator was a real-time audio/speech analyzer featuring time and frequency displays, accurate pitch estimation, accurate identification and visual representation of harmonic structures, and visually oriented vowel recognition using LP formant tracking. After fine-tuning the operation of the vowel recognizer for the main Portuguese vowels, it was quite disappointing to find that LP formant tracking was very reliable and robust when used with vowels uttered by adult male speakers, but not when used with vowels uttered by adult female or child speakers.

Our interest is motivated by two relevant application areas: language learning and speech therapy. These areas require interactive computer-based visual feedback of acoustic features extracted from phonemes, uttered in isolation, notably vowels, in order to reinforce or replace the natural acoustic feedback pathway. In the case of language learning, as pointed out by Palethorpe et al. (1996), it is known that the learning and training of vowels represents the first important stage in language learning, control of the phonetic system, and understanding of the vowel-consonant interactions. Therefore, a technology support for visual feedback of acoustic features reinforces the acoustic feedback and facilitates faster learning convergence to correct pronunciation. On the other hand, robust real-time phoneme recognition technology is also desirable in a clinical context involving speech therapy. The main advantages are that it increases the motivation of patients by providing a visual feedback environment, it provides a rich and stimulating diversity of practice exercises, facilitates their upgrade, and it accelerates rehabilitation programs by offering patients viable opportunities for self-training as a complement to formal speech therapy sessions conducted by a professional therapist.

In particular, children attending language learning or speech therapy programs represent the population segment that is most likely to benefit from phoneme recognition technology. In fact, children are naturally motivated to interact with computer-based games offering challenges, competition, and discovery. For example, a computer game may be designed using visual feedback of acoustic features and using sound as the preferred medium of interaction. Furthermore, the effectiveness of such a game may be increased by including automatic assessment functionalities.

Conceptually, these simple phoneme identification scenarios are just a small fraction of the broad automatic speech recognition problem, in the sense that they focus on robust feature extraction rather than on sophisticated language-aware classification strategies. Despite this fact, automatic identification of phonemes uttered in isolation suggests challenges in its own right since, to a large extent, human performance in recognizing isolated vowels (i.e., without context), in a speaker independent way, is not even approximated by current state-of-the-art technology. This perspective is shared by Dusan and Rabiner (2005), who support that there is "gap between machine and human performance of recognition systems across a wide range of problems." On the other hand, vowel recognition of child phonemes, which is the focus in this paper, has not received much research effort in recent years (Palethorpe et al., 1996).

B. The problem of vowel recognition at high pitch

According to Hess (1983), the fundamental frequency (or pitch) of the speech uttered by a human speaker can vary over a range of almost four octaves (50–800 Hz). In singing, including the *falsetto* register, the pitch may extend from 50 to 1800 Hz (Hess, 1983, p. 64).

If a vowel is whispered or is uttered at very low pitch, the corresponding magnitude spectrum is represented by a very dense structure of components that "illuminate" the envelope of the spectral representation with considerable detail. As a result, prominences (i.e., formants) in the spectral envelope that correspond to resonances of the vocal tract can be easily identified. A few spectral envelope models that are typical of the tonic Portuguese vowels /a/, /o/, /e/, /u/, and /i/, are illustrated in Fig. 1.

As formant estimation based on LP analysis tries to model the envelope of the magnitude spectrum of a sound, and its peaks in particular, reliable results are obtained when F0 is sufficiently low to resolve detailed prominences in the spectral envelope model such as those that are typical of the vowel /a/, as illustrated in Fig. 1. If F0 is sufficiently high as to not provide that resolution capability, as illustrated in Fig. 1 for vowel /i/, important spectral peaks are just "missed" and, as a result, formant estimates are not reliable. This scenario is even worse when F0 is comparable to the frequency of the first formant (F1), which happens frequently in singing, for example. Surprisingly, even in these cases, the HAS
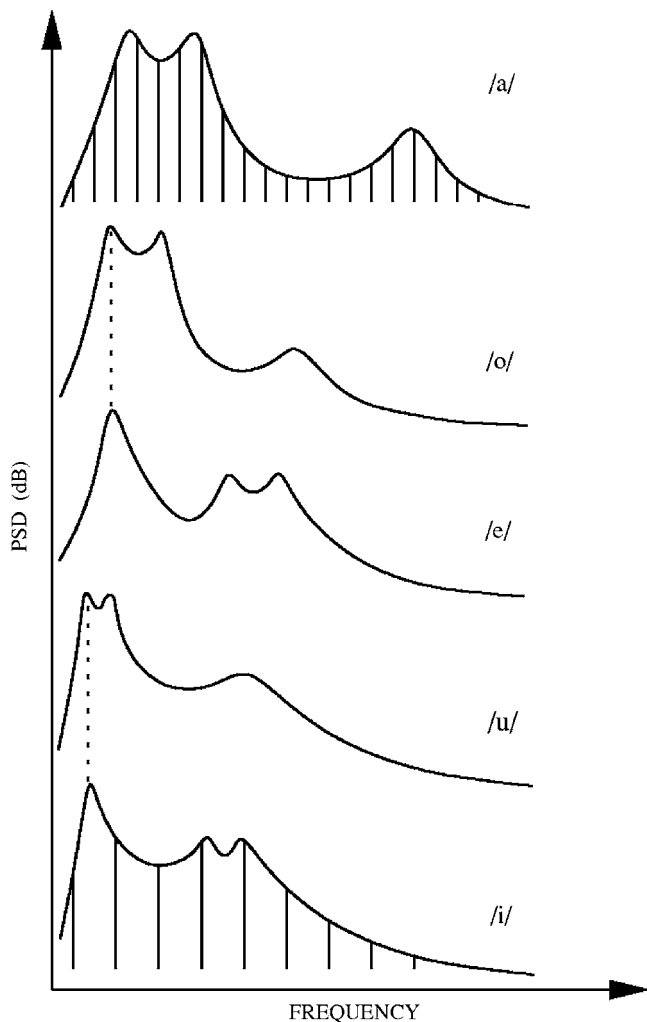
FIG. 1. Illustration of typical spectral envelope models pertaining to vowels /a/, /o/, /e/, /u/, and /i/ (the different models are shifted along the Y axis for convenience of visualization). When the pitch frequency is sufficiently low to resolve the prominences (i.e., formants) in the spectral envelope, as in the illustrative example at the top and regarding vowel /a/, LP analysis is able to provide reliable estimates of the formant frequencies. When the pitch frequency is relatively high, as illustrated at the bottom, formants are not well captured by the harmonic partials and as a consequence, LP analysis is not reliable in formant frequency estimation.

has an astonishing capability of recognizing vowels reliably, independent of the gender or age of the speaker, which suggests that the human auditory system is likely to use cues other than formants in the recognition process. An extensive characterization of the problem is discussed by de Cheveigné and Kawahara (1999).

C. Specific problems of formant estimation based on LP

It has been mentioned previously that formant tracking using linear prediction is a technique widely used to recognize vowel sounds. In particular, in the case of male adult speech, and for low sampling frequencies (e.g., 8 kHz), the representation in a plane of the frequencies of the first two formants estimated using LP allows a clear discrimination between the different vowels. This plane is normally known as the F1-F2 plane. As an example, Fig. 1 helps to illustrate

that the frequency locations of the first two formants (identified as local peaks in the spectral envelope) determine very strongly the vowel they are associated with. In particular, the second formant alone represents the main differentiating factor between vowels /o/ and /e/, and also between vowels /u/ and /i/.

However, the technique of formant estimation using LP is not speaker independent and the discrimination ability of the approach fails in the case of children speech or singing voice (Thorpe and Watson, 2000). In fact, as pointed out by Plomb (2002), "the overlap of the first two formants for different vowels due to men, women, and children is so considerable that the same configuration of formants may correspond with different vowels for the three, categories of speakers."

The most typical problems of LP-based formant estimation are as follows:

(1) The estimation accuracy depends strongly on the order of the LP model, the literature indicating that filter orders of practical interest are in the range 10–16.
(2) The sampling frequency of the speech signal affects significantly the operation of the LP filter and, in particular, when it is much higher than the frequency of the formants, their estimation is difficult and unreliable, even for high LP orders.
(3) The pitch of the voiced speech signal may act as an adverse distracting factor to the whole LP formant estimation process.

Some solutions commonly used to attenuate these problems and used to improve the performance of LP-based formant estimation, include:

(1) Decimation so as to reduce the sampling frequency and so as to separate low frequency formants,
(2) Preemphasis in order to facilitate LP estimation of high-frequency formants,
(3) Shifting the low-frequency poles of the LP model closer to the unity circle (of the Z plane) so as to emphasize different peaks in the spectral envelope,
(4) Nonlinear frequency scaling so as to increase the angular separation between formants, particularly at low frequencies [a popular technique is perceptual linear prediction, Honig et al. (2005)].

Due to the unreliability of formant estimation using LP techniques, most often, as described in Sec. I D, an automatic formant estimation procedure is followed by manual correction so as to eliminate errors and inconsistent results (Palethorpe et al., 1996).

D. Literature review

Several acoustic features that are alternative to formant tracking using LP have been proposed in the literature. Palethorpe et al. (1996) have suggested a technique based on critical band analysis and classification of speech signals. Their approach was motivated by the observation that inconsistencies exist in the relationship between a vowel classification based on formants and that based on listener judg-

Aníbal J. S. Ferreira: Static features in high-pitched vowel recognition   2391

ments. They compared the performance of formant tracking using 10-pole LP autocorrelation analysis, and spectral analysis on a critical band rate (Zwicker, 1961) up to 6400 Hz. Using both techniques and Gaussian classification, results have revealed that the best scores for correct classification of vowels were in both cases on the order of 60% for child speakers and about 90% for adult speakers. The authors have emphasized that although both critical band and formant based frequency analysis provide similar scores in the classification of vowels produced by both child and adult speakers, a fundamental difference lies in the fact that the LP-based formant tracking required a significant amount of hand-editing, while such hand-editing was not required in the case of the critical-band spectral analysis.[2]

Similar results have been presented by Zahorian and Jagharghi (1993), who compared global spectral-shape features (similar to cepstral coefficients) versus formants, on a data base of 30 speakers (10 men, 10 women, and 10 children). In a test of static vowel classification and using the Bayesian maximum likelihood classifier, results have indicated that a set of 10 global-spectral shape features scored about 77.0%, while a set of four formant features scored about 72%. The inclusion of the pitch as a feature had the effect to increase these scores slightly[3] (about 0.5% in the former case and about 3% in the latter case). The authors have noted that these results are far behind the human performance that has been concluded to be on the order of 91% correct identification. Using further tests, the authors have also concluded that duration cues appear to play a rather small role in vowel discrimination.

Thorpe and Watson (2000) have presented an alternative method that represents the vowel space directly by a factorial analysis of the harmonic amplitudes, without requiring explicit identification of formant frequencies. Taking into consideration that at high pitches, the vocal tract resonances are sampled only at multiples of the fundamental frequency, principal component analysis (PCA) is performed on the set of harmonics and significant eigenvalues are represented in bivariate scattergrams. As a reference, F1-F2 scattergrams are also obtained using a 12-order LP analysis and formant tracking. Vowels sung by professional singers at different pitches were recorded and analyzed. Results have shown that at high pitches, F1-F2 scattergrams lose the ability to distinctly represent the different vowels as they do at low pitches. PCA scattergrams show better results at high pitches but still a significant overlap between different vowel representations can be observed, which makes vowel discrimination difficult.

Hermansky et al. (1985) have proposed a vowel analysis procedure that is based on the perceptually based linear prediction model (PLP). This model performs a critical band analysis of the speech signal using 18 filters, preemphasis, intensity-loudness conversion, and inverse FFT delivering autocorrelation coefficients on the warped frequency scale. A fifth-order all-pole model is then used to extract at most two major peaks from the auditory spectrum. The PLP method identifies front vowels[4] by two strong peaks whose frequencies are very similar to F1-F2 formants, and back vowels[5] by a single peak whose center frequency and bandwidth denote

the spread of the F1-F2 formant region. Peak trajectories of vowels by male and female speakers were shown to be quite similar. Very convincing performance scores (approaching 98%) in speaker-independent automatic recognition of digit utterances are presented in a subsequent paper (Hermansky, 1990), but do not correspond to our target testing conditions since whole words are classified and thus dynamic time information has a significant influence on the results.

de Cheveigné and Kawahara (1999) suggest that vowel identification may be seen as a process of pattern recognition with missing data. Vowel recognition is implemented by evaluating the distance of a test signal to a vowel template using a weighting function derived after pitch estimation, and that emphasizes harmonic spectral regions much like a "harmonic sieve." The paper presents a rich discussion on alternative approaches addressing the problem of vowel recognition at high F0, and supports an F0-dependent smoothing not necessarily coincident with formant peaks. The proposed model of vowel recognition is presented in two versions (a frequency domain version and an autocorrelation version) and although the relative merits are discussed from a conceptual perspective, a complete evaluation of the performance of the proposed model is, however, not provided. As pointed out by Hillenbrand and Houde (2003), the missing data model of de Cheveigné and Kawahara was not evaluated on naturally spoken utterances. Only a few examples of template-matching are used to compare narrowband input spectra to smoothed spectral-shapes templates.

In a similar way, Mollis (2005) has addressed the relative performance of formant based-models and whole-spectrum models (including a method using excitation patterns and perceptual linear prediction cepstral coefficients) while using synthetic vowels in the testing methodologies. Results have indicated that representations that included only formant frequency information, performed poorly, although improvements could be obtained by adding relative formant magnitude information. The presented results are qualitative rather than quantitative because synthetic vowels with variations restricted to F2 and F3 formant frequencies have been used in the tests, and thus do not attempt to characterize a "practicable model of vowel perception," as admitted by the authors (Mollis, 2005, p. 1069).

Wet et al. (2004) depart from the observation that "there are still no tools available that can automatically extract exact true formants from speech reliably," and compare automatically extracted formant-like features ("robust formants" and HMM2) so as to compare them in vowel recognition tests to hand-labeled formants. Results have shown that the performance of automatically extracted formants was quite comparable to hand-labeled formants in a gender-dependent experiment, but was inferior in a gender-independent experiment. Furthermore, Mel-frequency cepstral coefficients (MFCC) features were also included in the tests and have exhibited the same or superior performance in vowel recognition tests (on the order of 91% correct recognition in gender-independent tests), especially under noisy acoustic conditions. The authors have therefore concluded that "for-

Aníbal J. S. Ferreira: Static features in high-pitched vowel recognition

mant representation of speech has no compelling advantages over representations that do not involve error-prone labeling decisions such as MFCCs."

The results presented by Wet *et al.* (2004) are not truly representative of speaker-independent tests since children utterances have deliberately been excluded from the test data with the argument that it "could not be guaranteed that the proposed formant extraction could handle children's speech appropriately" (Wet *et al.*, 2004, p. 1883). However, the authors support that previous knowledge on the gender of the speaker has a positive impact on the recognition performance.

Other recent papers on formant estimation do not address directly the problem of robust speaker-independent recognition of vowels but study other aspects such as the influence of noise (Bruce *et al.*, 2002; Chen and Loisou, 2004).

E. Structure of the paper

The rest of the paper is structured as follows. In Sec. II the PSC concept is described as well as the signal processing environment used to estimate PSC-related features and to assess experimentally its perceptual relevance. In Sec. III, three different feature sets and respective estimation methods are described that are used to benchmark the performance of PSC features. In Sec. IV the database of natural vowel sounds used to train and test the different vowel identification schemes is characterized. In Sec. V the classification results characterizing the relative performance of the four different feature sets in vowel identification tests are presented, and discuss their significance in Sec. VI. Section VII summarizes and concludes the paper.

II. THE PERCEPTUAL SPECTRAL CLUSTER CONCEPT AND FEATURES

The PSC concept has been inspired by Klatt's (1982) discussion regarding "prominent energy concentrations" in the magnitude spectrum of a vowel sound and a first experimental validation has been reported in Ferreira (2005). The PSC concept has emerged by realizing that in order for automatic sound identification to reach human performance, even in a task as simple as isolated vowel identification in real-time, attempts must be pursued to better understand and to get closer to the "signal processing in the human auditory system" (Rabiner and Juang, 1993, p. 132; Sroka and Braida, 2005, p. 421). In this spirit, the PSC concept is also influenced by Plomb's (2002) perspective that interpretation of a sound by the HAS is organized in stages, starting with raw features such as sound intensity (and the resulting sensation of loudness), and progressing to more elaborate stages of perception corresponding to pitch and timbre for example, by incorporating more features (or more details of the same features), until a complete psychoacoustic percept[6] is formed. This perspective is suggested in Fig. 2. Ultimately, not only the recognition of a specific sound is achieved at the linguistic level (such as the identity of a spoken vowel), but an interpretation is also established regarding the quality of the vowel and the identity of the speaker, by using additional features. These are related, for example, to the glottal pulse
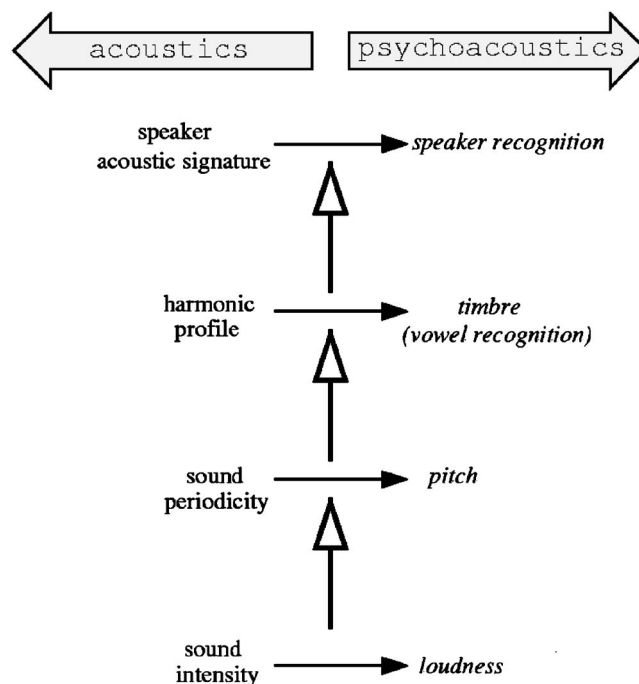

FIG. 2. Conceptual approach to sound interpretation. As more elaborate features are captured from an acoustic stimulus, the more complete the psychoacoustic interpretation is of the acoustic event.

shape, or to the detailed and persistent spectral resonances that correspond to the acoustic signature of a specific individual speaker. As also argued in Plomb (2002), "speaker-specific differences in pitch and timbre, as well as timing cues, provide the basis for separating voices into [perceptual] streams."

The PSC concept is strongly rooted in the idea that the human recognition of a sustained *voiced* vowel results from both the identification of its pitch, and timbre. Timbre is commonly seen as the "color" of a sound and, in the case of a harmonic sound such as a vowel utterance, it is a function of the spectral power of its partials. Thus, for a voiced vowel sound, timbre analysis requires the identification of the underlying harmonic structure.

The perception of a harmonic structure and its attributes plays a very important role in the auditory ability to recognize and segregate sounds. In fact, as argued by Plomb (2002), "in listening to two or more simultaneous complex tones-human voices or musical voices-, the auditory system is continuously 'testing' which sinusoidal components are harmonics of the same fundamental." As a result, partials that are harmonic of the same fundamental frequency (F0) are fused or integrated on a single auditory object (or acoustic percept). On one hand, this helps to segregate among harmonic structures that have a different pitch.[7] On the other hand, the perceptual aggregation of harmonic partials also helps to differentiate among harmonic structures possessing a similar pitch, due to differences in timbre. This fact is invoked by the American Standards Association as a major aspect in defining timbre as a perceptual quality of a sound, although it is admitted that the intensity and temporal characteristics of the sound also have an influence.
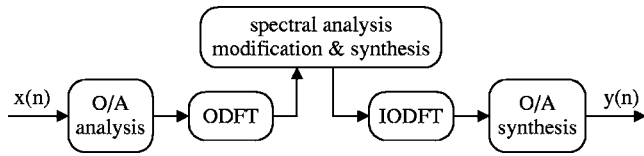
FIG. 3. Signal processing environment used to test the PSC concept.

The PSC concept builds on this perceptual integration of partials pertaining to the same harmonic structure, and tries to identify clusters of harmonic partials and their attributes, which explain the ability of the HAS to differentiate among vowels. In other words, inasmuch as the different partials of a harmonic structure are fused in a perceptual sensation of pitch (Moore, 1989), it is admitted that a second level of perceptual integration involving the harmonic partials within each PSC is carried out by the human auditory system, giving rise to the sensation of spectral color and meaning. This hypothesis has been tested extensively using the signal processing environment illustrated in Fig. 3. This environment is essentially a subset of processing tools pertaining to an audio coder (Ferreira, 1996a), allowing analysis, modification, and synthesis of sounds in the frequency domain. The main processing blocks involve the following:

(1) Segmentation of the input audio signal in segments $x(n)$ of length N, with 50% overlap, and windowing of each segment with a window corresponding to the square root of a shifted Hanning window (Ferreira, 1998):

$$h(n) = \sin \frac{\pi}{N}\left(n + \frac{1}{2}\right), \quad n = 0, 1, \ldots, N-1, \qquad (1)$$

(2) Time-to-frequency transformation by means of an odd DFT, defined as

$$X_{\text{ODFT}}(k) = \sum_{n=0}^{N-1} h(n)x(n)e^{-j(2\pi/N)(k+1/2)n},$$

$$k = 0, 1, \ldots, N-1, \qquad (2)$$

(3) Frequency-to-time transformation by means of an inverse odd-DFT transformation,
(4) Windowing of the reconstructed segment with the same window used for analysis, and signal reconstruction using overlap-and-add with 50% overlap.

If there is no spectral modification, this analysis/synthesis environment is perfect reconstructing (Vaidyanathan, 1993), i.e., $y(n)=x(n\neg n_0)$, where n_0 represents a delay. The processing environment illustrated in Fig. 3 also has the capability of identifying sinusoidal components pertaining to an existing harmonic structure (Ferreira, 1996b), and modifying them with regard to frequency, magnitude, or phase (Ferreira, 2001). This processing is implemented in the ODFT frequency domain. Although it delivers essentially the same information as the DFT, the ODFT is more convenient from the point of view of signal analysis and modification (Ferreira, 1998).

Using this signal processing environment with $N=1024$ samples and 32 kHz sampling frequency, we could manipu-

late the characteristics of individual groups of partials pertaining to the harmonic structure of a voiced vowel, and assess what specific changes led to a loss of identity of the vowel, or led to a vowel conversion. Although some of the conclusions of these experiments are addressed in Sec. VI, for the purpose of defining the PSC concept and explaining what features have been extracted based on this concept, here an illustration is given, using one example of how the concept and identification of a PSC has emerged.

Figure 4 represents the spectrogram of short audio segments extracted from the stationary region of voiced vowel sounds uttered by a young female. This signal is included in the database. Figure 4 also represents the short-time power spectral density (PSD) of the individual vowel sounds. For each vowel sound, different harmonic partials have been manipulated so as to study their impact on vowel identity. Using many vowel sounds from different speakers and after extensive testing, it has been concluded that only two clusters of adjacent harmonic partials were necessary to preserve vowel identity, the remaining ones being either masked or not relevant in contributing to the vowel identity. As a result, the decision was made to name those clusters PSC1 and PSC2, PSC1 being on the left of PSC2. Figure 5 illustrates the two clusters of harmonic partials above the noise floor (PSC1 and PSC2), which resulted from the natural vowel sounds illustrated in Fig. 4, as a result of the manipulation just described. These manipulations also included a magnitude smoothing with the purpose of avoiding small local peaks among adjacent partials. In addition, the pitch frequency is normalized to the average pitch of the speaker in order to focus the identification process on static cues.

Repeating this manual procedure for each vowel sound in the database (to be described in Sec. IV) has proved to be a tremendously time-consuming task. In order to avoid fragile judgments and results, a criterion has been adopted that the spectral modification emphasizing PSC clusters should not only preserve vowel identity but also a significant degree of speaker identity. The insight gained by performing these spectral modifications led to the design of an automatic PSC estimation algorithm described in Sec. II A.

A. Estimation of PSCs

A signal processing algorithm was developed to automatically detect PSCs, on a frame-by-frame basis, by identifying their lower and upper borders and average spectral power. The main processing steps[8] of the algorithm can be structured into two major blocks (PSC preprocessing and PSC merge) and are as follows:

(1) A new frequency domain is created that includes all harmonic partials in the magnitude spectrum of the voiced vowel (i.e., the noise floor between two adjacent harmonic partials is ignored),
(2) A magnitude smoothing in the new frequency domain is implemented so as to avoid small local peaks,
(3) All local peaks are identified as potential PSC candidates,

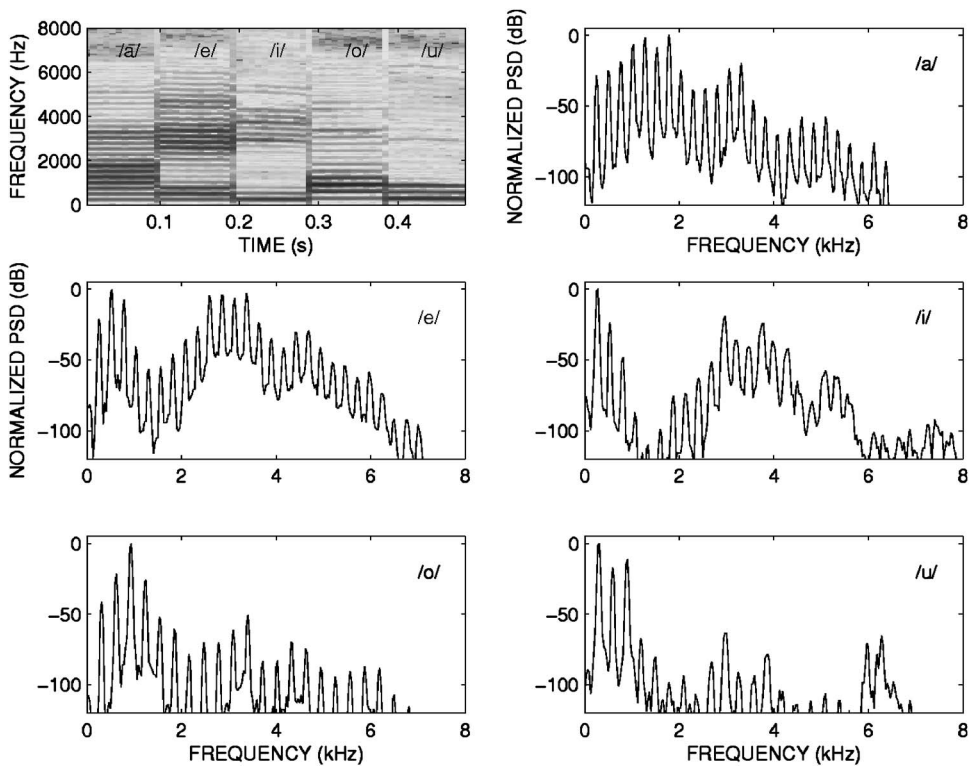Aníbal J. S. Ferreira: Static features in high-pitched vowel recognition

FIG. 4. Spectrogram and short-time PSD representations pertaining to five vowels uttered by a 13 year old young female.

(4) Starting from the center of each PSC candidate, left and right borders are found by absorbing into the PSC neighboring partials whose magnitude is not below 8 dB[9] of the average magnitude of the PSC (this average magnitude is updated every time one more partial is absorbed into the PSC),

(5) The above PSC preprocessing does not merge different PSCs, but may result in PSCs with abutting borders corresponding to local minima; these PSCs are first identified and, if their absolute magnitude difference is below 8 dB, a PSC decimation is carried out by merging PSCs,

(6) Adjacent but nonabutting PSCs are identified and, if sufficiently close to each other, their magnitude difference is evaluated and eventually are merged if the magnitude criterion of the previous step is met.
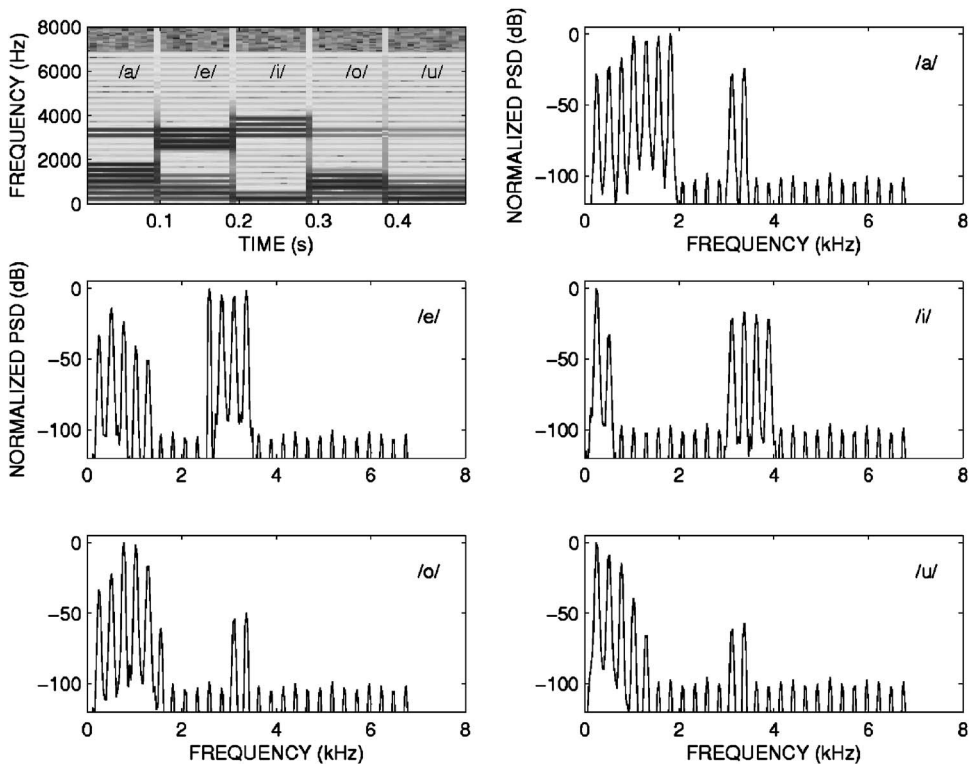
FIG. 5. Spectrogram and short-time PSD representations of resynthesized vowel sounds resulting from manipulation on the harmonic partials of the original sounds illustrated in Fig. 4, in such as way as to preserve vowel identity. Two spectral clusters of harmonic partials were concluded to be sufficient for clear recognition upon listening.
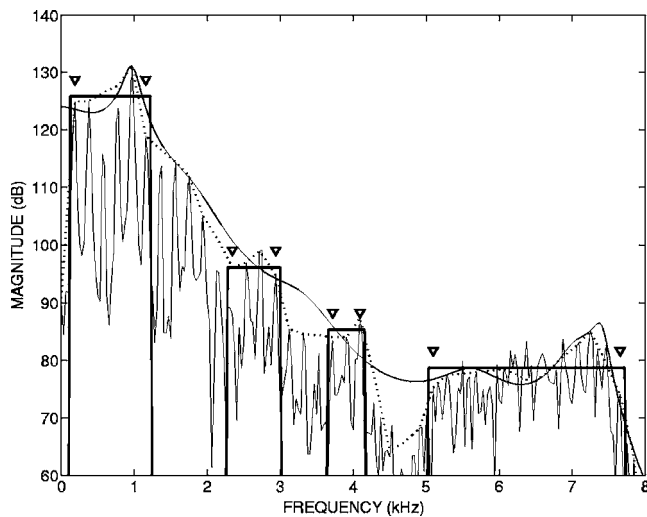
FIG. 6. Short-time PSD of a sound frame corresponding to the utterance of vowel /a/ by a female speaker (thin solid line), spectral envelope derived from the magnitudes of the harmonic partials and after smoothing (dotted line), spectral envelope model derived from linear prediction analysis with order 16 (smooth solid line), and identification of the borders (triangles) and magnitude of the PSCs found (thick solid line).

This algorithm is iterated for each frame until there are no more PSCs to merge. After execution of the algorithm, a mapping to the original frequency domain is performed of the boundaries and average magnitude of all PSCs found. This allows the verification of the correctness of the processing both in a graphical and auditory perspective.

As one example, Fig. 6 depicts the result of the algorithm for one frame of the sound corresponding to vowel /a/ as uttered by a female speaker. Figure 6 shows the short-time PSD of the original signal, the smooth spectral envelope model based on the magnitudes of all harmonic partials, the spectral envelope model derived from a 16th-order linear prediction analysis, and the borders and average magnitude of all PSCs found.

In order to test the perceptual impact of the information delivered by the algorithm, the processing environment illustrated in Fig. 3 is used in order to resynthesize the vowel signal after spectral modification. This modification preserves only harmonic partials within the borders of any selection of PSCs, and may either preserve the original magnitudes or enforce magnitude equalization by the average magnitude of each PSC. In the resynthesis, the original initial phases of the preserved harmonic partials may be used or replaced by a random phase value (in which case, phase continuity is enforced in subsequent frames).

B. Extraction of PSC features

The algorithm described in Sec. II A has been fine-tuned and tested with most of the vowel recordings included in the database (which is described in Sec. IV). The resynthesis and listening of the modified vowel spectra led us the following main conclusions:

(1) The two PSCs with highest average magnitude are sufficient for clear vowel identification.
(2) A modest degree of magnitude smoothing on the preserved harmonic partials does not significantly affect either vowel recognition or speaker identification.
(3) The replacement of the original initial phase of the preserved harmonic partials by a random phase value (and ensuring subsequent proper phase continuity) appears to interfere with speaker identity.
(4) The PSC falling below about 2 kHz has the highest impact on vowel identity and, in particular, the suppression (or strong attenuation) of a single harmonic partial may transform a vowel sound into another (for example, an /a/ may be transformed into an /o/).
(5) Harmonic partials at the left border of the PSC falling below about 2 kHz do not contribute significantly to vowel identity, their impact is more related to quality of the vowel and "presence" of the vowel sound (which is consistent with the well-known phenomenon of missing fundamental).
(6) The specific shape of the spectral envelope described by the magnitudes of the harmonic partials within the PSC falling below 2 kHz can be quite diverse, from convex, flat, until concave, and it impacts mainly on vowel quality and speaker identification rather than on vowel identification.
(7) The feature pertaining to the PSC falling above about 2 kHz and with highest impact on vowel identity is its average magnitude; as an example, if this magnitude is manipulated so as to be significantly attenuated, an /e/ sound is transformed into an /o/ sound, and an /i/ sound is transformed into an /u/ sound.
(8) Very realistic vowel transformations can be obtained by using the shape of the spectral envelope described by the magnitudes of the harmonic partials pertaining to one vowel, in order to modulate the magnitudes of the harmonic partials pertaining to a different vowel.

These conclusions, with the preliminary results reported in Ferreira (2005), and experiments with real data (i.e., with audio sounds) suggested that after the automatic identification of the two most relevant PSCs, as described earlier, a feature vector including as few as five features should be able to provide good classification results. The chosen features are:

(1) Pitch frequency (F0),
(2) Center of gravity of PSC1, designated as COG1,
(3) Center of gravity of PSC2, designated as COG2,
(4) Right border of PSC1, designated as PSCR,
(5) Decibel difference between the average magnitude of PSC1 and that of PSC2, designated as difdB.

If ω_L and ω_R are, respectively, the frequencies of the left and right borders of a PSC, with L and R integers and $L \leq R$, the COG frequency is obtained as

Aníbal J. S. Ferreira: Static features in high-pitched vowel recognition

$$\mathrm{COG} = \frac{\sum\limits_{k=L}^{R} \omega_k |X_{\mathrm{ODFT}}(\omega_k)|^2}{\sum\limits_{k=L}^{R} |X_{\mathrm{ODFT}}(\omega_k)|^2}. \qquad (3)$$

It should be noted that the definition of "center of gravity" given here differs significantly from the definition given by Chistovich and Lublinskaja (1979) and other authors, to the same concept, although there are some aspects in common. In fact, Chistovich and Lublinskaja (1979) and other authors support that in the case of back vowels, whose first two formants (F1 and F2) are typically very close to each other, the HAS does not perceive the two formants separately, but performs instead a spectral integration spanning a frequency range of about 3.5 bark (or about 350 Hz at low frequencies). Thus, stimuli with formants closer than this limit are found to be perceptually equivalent to one peak stimulus, with the peak position determined by the center gravity of the original two peaks (Lass, 1996, p. 489). To a large extent, this is consistent with our experimental findings (this equivalence does not impact on vowel identity but impacts on signal quality), except that the bandwidth of each PSC is not constrained to be 3.5 barks (as considered by Chistovich and Lublinskaja, 1979), but results instead from a magnitude criterion that starts from the center of gravity of each PSC candidate, and captures neighboring harmonic partials according to a magnitude threshold. As a result, a PSC may span a frequency range wider than 3.5 barks (or even less than this; in fact, some PSCs have been found to be as narrow as a single harmonic partial).

For each audio frame, a feature vector is therefore obtained as

$$[\mathbf{v}_{\mathrm{PSC}}]^T = [\mathrm{F0}, \mathrm{COG1}, \mathrm{PSCR}, \mathrm{COG2}, \mathrm{difdB}]. \qquad (4)$$

Figure 7 illustrates the main processing blocks involved in obtaining a PSC feature vector as described in this section. Figure 7 also illustrates the computation of alternative features that will be compared to PSC features in automatic vowel classification tests. The alternative features consist of the frequencies of the first four roots of the LP polynomial that is obtained using the Levinson-Durbin recursion algorithm. The autocorrelation coefficients are obtained using a smooth spectral envelope model (based on the magnitude of the harmonic partials), and by computing an inverse odd-DFT transform (i.e., by taking advantage of the Wiener-Khintchine theorem[10]). Thus, these alternative features represent formant-like features and will be addressed in Sec. III.

III. REFERENCE METHODS OF FEATURE EXTRACTION

In order to evaluate the relative performance of PSC features in classifying high pitched vowel sounds, three alternative techniques of feature extraction are used: plain LP analysis and polynomial root analysis, modified LP analysis and polynomial root analysis as described in Sec. II, and Mel-frequency cepstral coefficients (MFCC) (Davis and Mermelstein, 1980). All techniques were configured to pro-
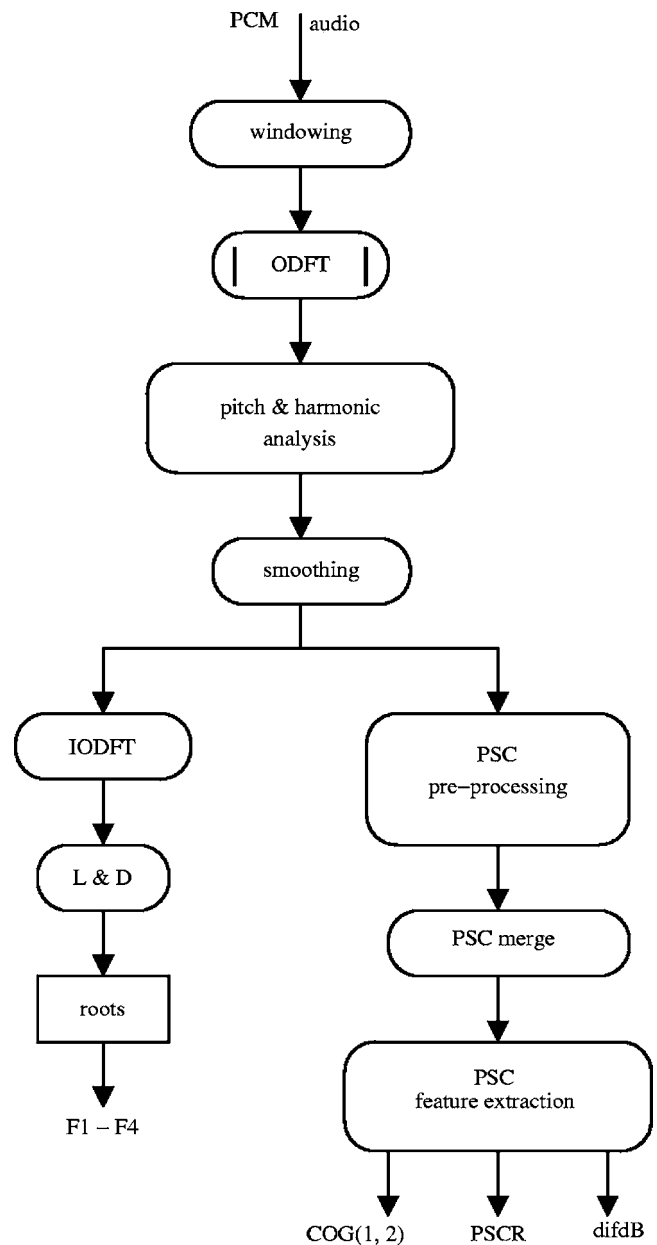

FIG. 7. Signal processing algorithm used to estimate PSC features and modified linear prediction features.

cess 1024-sample long audio frames, and presume that the sampling frequency is 32 kHz. In the following each technique is briefly characterized.

LPC: The plain LP and polynomial root analysis technique corresponds to the classic technique of formant estimation by finding the poles of the LP filter that models the vocal tract resonances. In our implementation, its main processing steps correspond to (32 ms) Hamming windowing, autocorrelation computation, determination of the coefficients of the all-pole LP filter using the Levinson-Durbin recursion, polynomial root analysis delivering the poles of the LP filter, determination of the angular position of the first four poles (excluding those over the real axis of the Z plane), and denormalization of the angular positions to the Hertz scale, which delivers coefficients F1, F2, F3, and F4. These coefficients correspond roughly to formants, although a strict

correspondence to the formant concept would require a significant amount of hand-labeling as explained in Sec. I C. Since in this paper an evaluation is desired of fully automatic feature extraction techniques that are amenable to real-time processing, the feature vector of the LP coefficients (LPC) is formed using the above-mentioned F1–F4 values in addition to the pitch frequency, i.e.,

$$[\mathbf{v}_{\mathrm{LPC}}]^T = [F0, F1, F2, F3, F4]. \tag{5}$$

LPC-WK: Another set of four formant-like coefficients are obtained using the algorithm depicted on the left-hand side of Fig. 7 and described in Sec. II B. To a large extent, this algorithm prevents "formants" from locking to the harmonic frequencies and therefore should not replicate the estimation errors of the plain LP technique, as explained in Sec. I. As this technique uses a smooth spectral envelope model and takes advantage of the Wiener-Khintchine theorem, it is conveniently designated as LPC-WK. Although the resulting F1, F2, F3, and F4 coefficient values are different from those of the previous technique, its feature vector is formed in a similar way (also including F0):

$$[\mathbf{v}_{\mathrm{LPC\text{-}WK}}]^T = [F0, F1, F2, F3, F4]. \tag{6}$$

MFCC: MFCC coefficients (Davis and Mermelstein, 1980) are used in almost every state-of-the-art speech recognition system, and therefore represent well established and widely accepted efficient features, despite the fact that their psychophysical interpretation is not obvious. In computing the MFCC coefficients, Slaney's Auditory Toolbox[11] is used after adjusting the MATLAB code for 32 kHz sampling rate processing and 1024-samples FFT analysis. The MFCC feature vector is formed using the 16 MFCC coefficients following the energy coefficient (i.e., c_0):

$$[\mathbf{v}_{\mathrm{MFCC}}]^T = [c_1, c_2, \ldots, c_{16}]. \tag{7}$$

IV. TRAINING AND TESTING DATABASE

Our motivation in this research is to find static features whose performance in automatic identification of vowels uttered in isolation, approaches human performance, in a speaker-independent way. In addition, the duration of each vowel in the database should be as short as possible while maintaining robust recognition by a human listener (after proper fading-in and fading-out of the time signal so that artifacts such as clicks are not audible). A short duration also has the advantage of making it easier to capture a short audio segment in a stationary region of the recorded vowel utterance, where only static features are significant.

Although the shortest time needed for proper recognition of an isolated vowel by a human listener may vary from vowel to vowel, a rather conservative duration for all vowel segments in the database is adopted: 100 ms. In informal verification tests, in which the first 10 ms and the last 10 ms of the 100 ms duration of a vowel signal have been smoothed so as to avoid clicks, it has been confirmed that good identification based on listening was not affected, making it safe to presume 100% correct identification by a human listener.
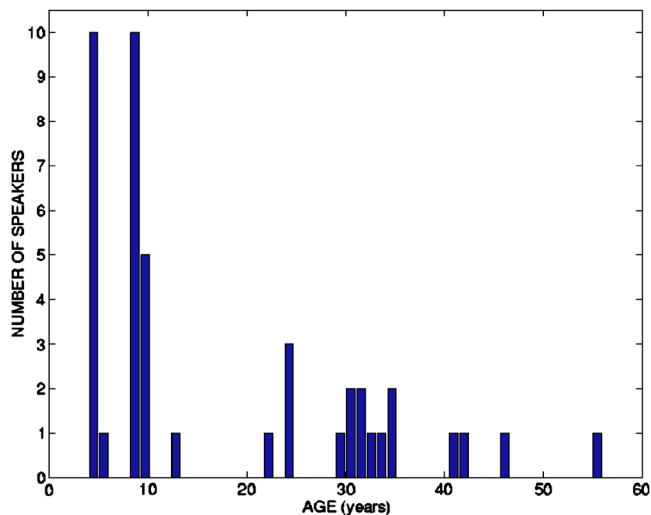

FIG. 8. (Color online) Age distribution of the speakers participating in the recordings.

Given that the focus is on vowel recognition of natural utterances at high pitch, for the recordings volunteer speakers (mainly child and female speakers) have been recruited from a kindergarten school, an elementary school, and an university school. In the first two cases, the educators of the prospective volunteer child speakers have been informed about the purpose of the recordings and associated research, and obtained their written permission before the recording session. In total, 27 child speakers, 11 adult female speakers, and 6 adult male speakers participated in the recordings. Figure 8 shows the age distribution of the speakers and it can be concluded that the highest density corresponds to 5 and 9 year old speakers.

Each speaker was asked to utter in sequence and in a sustained way, the most common tonic Portuguese vowels: /a/, /e/, /i/, /o/, /u/. Instructions (by way of example and rehearsal) were given so that the duration of each vowel was 1 to 2 s. Particularly with child speakers, in order to make them feel more comfortable and produce more natural (i.e., not stressed) sounds, at most three trials were done before the effective recording. The voice utterances were captured in a quiet room using a laptop (Compaq-Presario 1500), an electret microphone with good frequency response between 50 and 8000 Hz, and an audio editor that was configured for 32 kHz sampling frequency.

After the recordings, the vowel sounds were edited and the most stable 100-ms-long region was manually segmented, labeled, and entered into the database. The decision regarding stability was taken by manually selecting a region of the recorded vowel with stable spectrogram, time plot, and sound. A total of 220 (=44 speakers × 5 vowels) labeled sounds are included in the database.[12] This database was used for both training and testing of the four different feature sets and classification methods (PSC, LPC, LPC-WK, and MFCC), as explained in Sec. V.

A preliminary analysis on the pitch of all speakers was performed in order to verify the robustness of the pitch estimation algorithm. This algorithm is a frequency domain pitch estimator that is inspired by Goldstein's (1973) results.
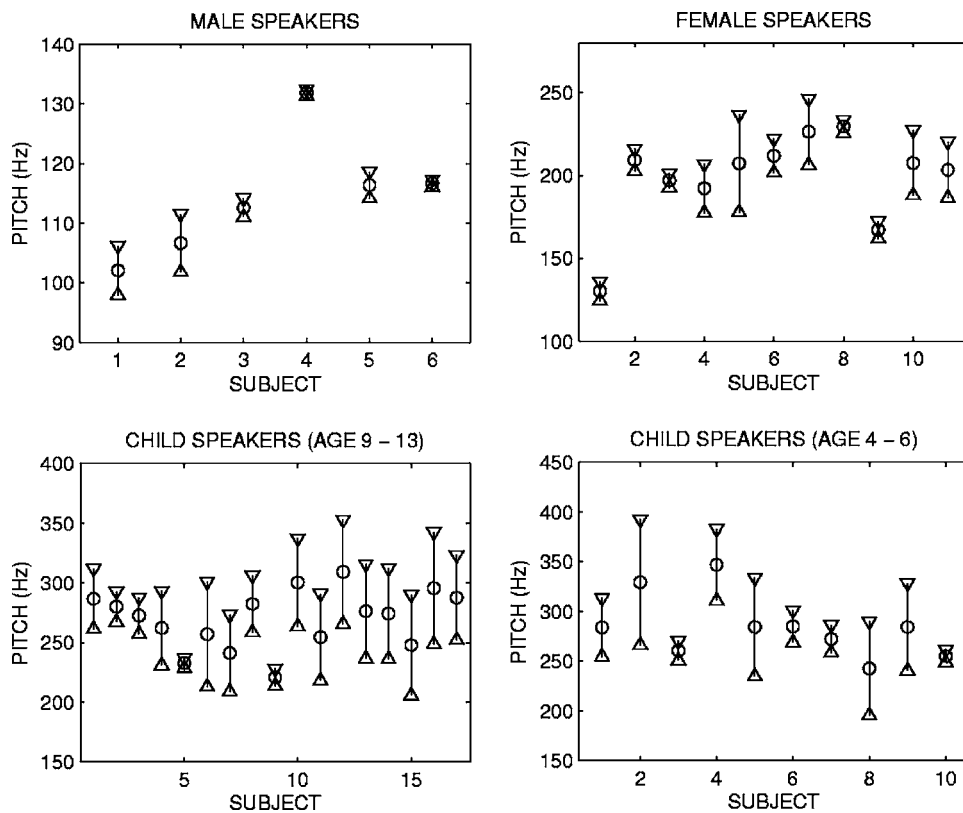
Aníbal J. S. Ferreira: Static features in high-pitched vowel recognition

FIG. 9. Statistical analysis (mean±standard deviation) of the pitch of all speakers. For each speaker, the statistical analysis includes the five vowel sounds in the database.

It is also described in Hess (1983, p. 425), and it takes into account the specificity of the odd-DFT and analysis window (Ferreira and Sinha, 2005). Furthermore, it also takes into consideration (and therefore is robust to) the fact that the harmonic structure of a voiced sound is likely to exhibit missing fundamental or discontinuities.

The statistical results are depicted in Fig. 9 for all speakers arranged into four categories. For each speaker, the mean and standard deviation were computed using the pitch automatically estimated for all frames over all five vowel sounds. As each vowel sound in the database is 100 ms long, given that the sampling frequency is 32 kHz, and that the analysis is carried out on 1024-sample frames with 50% overlap; pitch and other features are collected for 5 frames in each vowel. This represents a total of 25 pitch values over all five vowels, for the same speaker. Results in Fig. 9 indicate that pitch ranges are consistent with the expected values for the different categories of speakers, and that some speakers were very "tuned" (i.e., showed little pitch variance), while others exhibited a considerable pitch variance. A more detailed analysis of these cases has revealed that the variance is explained due to intervowel pitch differences and not due to intravowel pitch differences.

V. CLASSIFICATION RESULTS

Each feature vector is modeled as a multivariate Gaussian random variable with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$. The statistical behavior of each PSC feature has been observed in order to anticipate (and better interpret) its individual discrimination capability. As an example, Fig. 10 il-

lustrates the scattergrams of the PSC features for vowel /a/, as a result from the analysis of the complete database (which corresponds to 44 speakers × 5 frames/vowel=220 tokens). Each feature scattergram is represented as a function of the pitch frequency. It can be seen that features COG1 and COG2 exhibit a clear dependency on F0, which is denoted by the slope of the illustrated lines that best fit the data in the least-squares sense. This is consistent with a known similar effect regarding formants (Maurer et al., 1992; Rabiner and Juang, 1993). In this regard, it is interesting to note that previous studies with synthesized vowels have revealed vowel identification by humans is less reliable when F0 is high, in which case the transfer function of the vocal tract is sparsely sampled in frequency[13] (Diehl et al., 1996; Ryalls and Lieberman, 1982).

The dependency of PSCR on F0 is not as well evident because of the peculiar representation of its scattergram. In fact, because the right border of each PSC corresponds to a harmonic partial, the scattergram plane is "sampled" according to lines following the rule $n_R F_0$, where F_0 is the pitch frequency and n_R represents the order of the partial on the right border of the PSC. The fact that lines are clearly visible is also evidence that the pitch estimation algorithm is accurate and robust. The difdB feature does not exhibit any relevant dependency on F0 for any vowel.

Another perspective can be appreciated by looking at the Gaussian models pertaining to the same PSC feature for all five vowels. As an example, Fig. 11 represents those models for COG1 data. Despite some significant overlap, there are distinctive trends (which were followed in sketching Fig. 1) that are consistent with spectral observations.
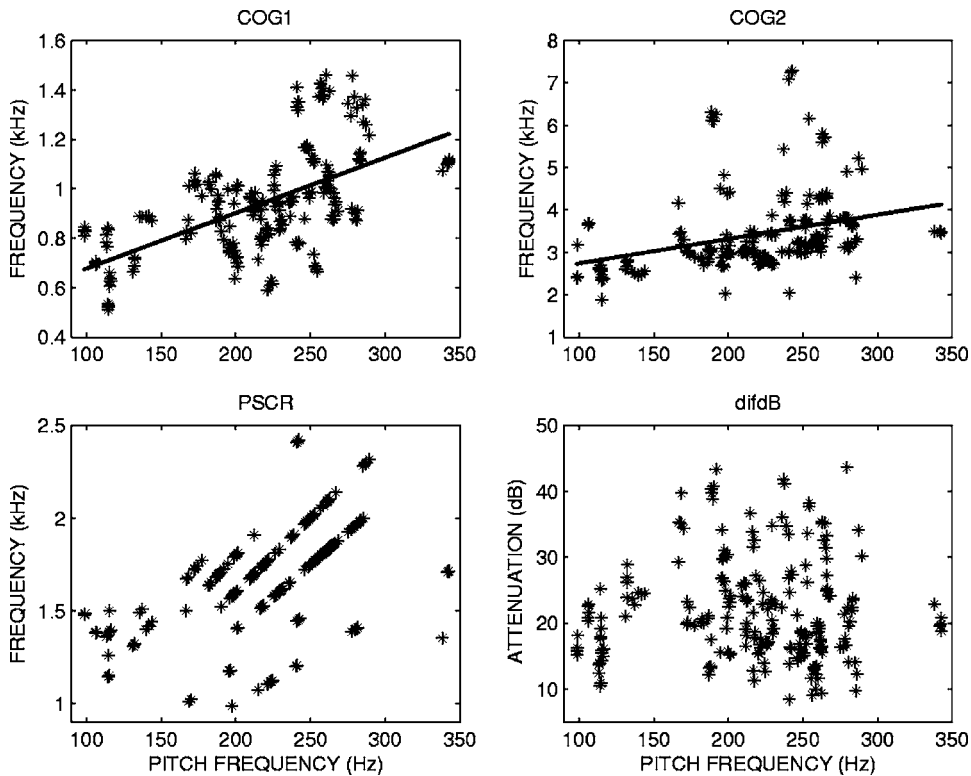
FIG. 10. Scattergrams of the selected four PSC features regarding vowel /a/, as a function of the pitch frequency. The lines in the COG1 and COG2 scattergrams represent the linear models that best fit the data in the least-squares sense.

A. Classification criterion and test procedure

Our classification criterion is based on the Mahalanobis distance, defined as

$$d^2 = (\mathbf{v} - \boldsymbol{\mu})^{\mathrm{T}} \mathbf{C}^{-1} (\mathbf{v} - \boldsymbol{\mu}). \tag{8}$$

In this equation $\mathbf{v}$ represents the multivariate feature vector (test vector), and $\boldsymbol{\mu}$ and $\mathbf{C}$ represent, respectively, the mean and covariance matrix computed from the training data for a specific vowel. The Mahalanobis distance is very convenient because it takes into account the correlation between variables of the data vector, it automatically ensures appropriate scaling of each variable, and it is reasonably amenable to real-time implementation. Furthermore, by forcing $\mathbf{C}$ to be diagonal, it becomes very easy to evaluate the impact of

FIG. 11. Gaussian PDF models concerning COG1 data for all five vowels.

ignoring the correlation among variables of the feature vector, in which case the Mahalanobis distance reduces to the normalized Euclidean distance. The classification of a test vector results from the evaluation of the Mahalanobis distance to all vowel templates, and the identification is declared for the vowel which minimizes the distance.

Classification results have been obtained using a K-fold cross-validation procedure (Flexer, 2006). According to this procedure, the training data and test data are always mutually exclusive. In each run, about 90% of the data is used for training and the remaining 10% is used for testing. Accordingly, in each run of the classifier, the utterances of 40 speakers (about 0.9×44) were used to build vowel templates, while the utterances of the remaining 4 speakers were used for vowel classification. This procedure was repeated 11 times so that the utterances of a given speaker were included in one test data set. Final classification results were obtained by evaluating the mean, confidence interval, and analysis of variance (ANOVA) of the results of the 11 runs.

B. Classification performance

The classification performance of PSC features has been evaluated on an individual basis, and also when all PSC features are used together. In the first case, the objective was to rank PSC features according to its discrimination capability, and in the second case the objective was to assess the impact of ignoring correlations between the different features (by forcing a covariance matrix to be diagonal). The results are displayed in Fig. 12. Three main conclusions can be extracted from Fig. 12. First, the PSCR feature (i.e., the right border of PSC1) exhibits the best individual discrimination capability, which may appear to be a surprising result. However, this is totally in line with observations derived from our
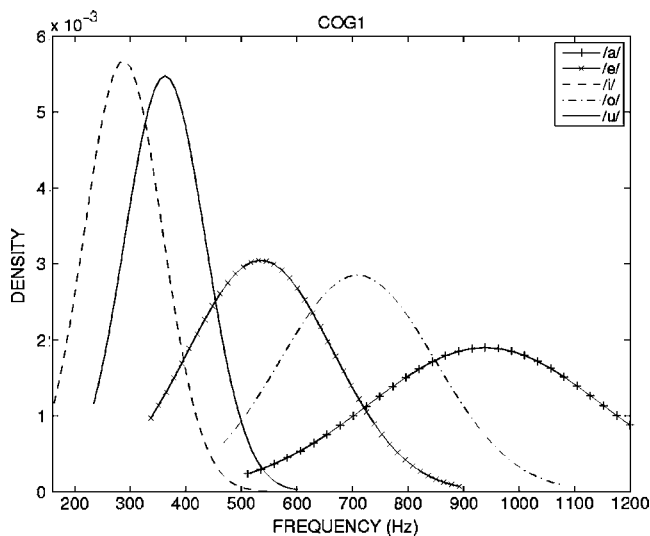
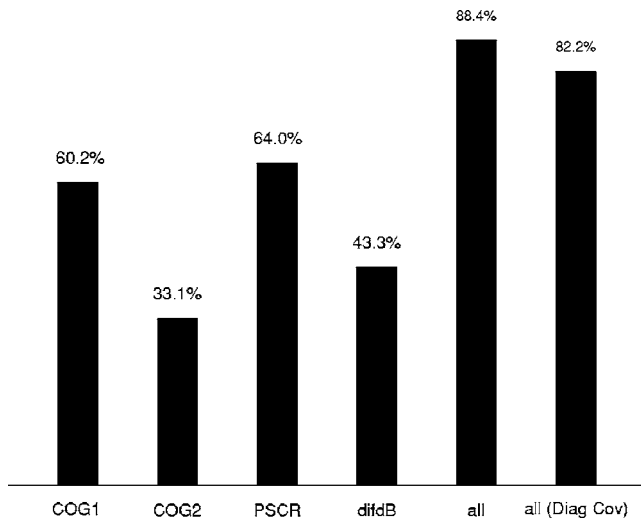Aníbal J. S. Ferreira: Static features in high-pitched vowel recognition

FIG. 12. Overall classification scores for individual features (COG1, COG2, PSCR, and difdB) and for the joint feature vector. In this case, the impact of diagonalizing the covariance matrix is also assessed.

experimental tests that indicate that the right border of the first PSC has a dominant perceptual effect, frequently "overriding" other cues such as the magnitude difference between the first two PSCs (i.e., the difdB feature). Second, the COG2 feature appears to have the lowest individual discrimination capability, which is also in line with evidence resulting from our experimental tests in resynthesizing spectrally modified vowel sounds. Essentially, we have concluded that as long as an appropriate magnitude was assigned to the second PSC, the correct (perceptual) vowel impression was obtained, independent of the fact that COG2 was shifted to the left or to the right by a few hundred hertz. Lastly, the middle-rank position of difdB feature is consistent with the fact that it plays a significant discrimination role only in specific cases, such as in discriminating /o/ from /e/, and in discriminating /u/ from /i/. In case of vowel /a/ for example, the difdB feature appears to be perceptually overridden by the PSCR feature.

The performance when all PSC features are used is 88.4% correct identification, and there is a penalty of around 6% if the covariance matrix is diagonalized. This indicates that the PSC features are not mutually independent. This result does not come as a surprise since it could be easily anticipated that there are clear correlations between PSC features; a notable example is the feature pair COG1-PSCR.

Table I shows the average correct identification results for all tested feature sets and associated feature extraction procedures. The 95% confidence intervals of the true means are also indicated as a way to express the uncertainty of the

cross-validation procedure. In Table I, the results for each feature set are presented in two versions, a version that uses the full covariance matrix in the Mahalanobis distance, and another version constraining the covariance matrix to be diagonal. As expected, in the latter case, scores are worse than those obtained in the former case, the largest difference (about 10%) being observed when LPC features are used, which indicates that the first four roots of the LP polynomial exhibit a substantial cross correlation.

Results in Table I show that while the lowest performance is achieved with LPC features, the best performance is achieved with MFCC features. LPC-WK features and PSC features exhibit similar scores and a paired t-test has therefore been implemented in order to assess the significance of the difference between the means. It was concluded that taking the cross validation results when both methods use full covariance matrices, $|t| = |0.1182| < t_{95\%, \text{Dof}=10} = 2.23$, which indicates that according to a two-tailed t-test at 95% confidence level and with 10 degrees of freedom (Dof), the difference between the means $(88.7-88.4)$ is not significant. Similarly, taking the cross-validation results when both methods use diagonal matrices, $|t| = |0.4148| < t_{95\%, \text{Dof}=10} = 2.23$, which again indicates that the means do not differ significantly. Therefore, PSC and LPC-WK features exhibit a comparable performance and are effectively able to capture highly discriminative information. This helps to confirm that when pitch information is taken into consideration to smooth the vowel spectrum, prior to LP analysis and root extraction, a more stable spectral envelope is obtained that gives rise to better identification scores. An explanation is that smoothing avoids the "harmonic locking" problem. In general, these results show that to a statistically significant extent, the goal of finding features that are alternative to LP analysis and formant estimation, and that are able to capture discriminative perceptual cues, has been achieved. On the other hand, results also suggest that there are structural similarities between the two methods as will be addressed in Sec. VI.

Results regarding MFCC features confirm that they are able to capture discriminative static cues more effectively than all other tested features. An argument could be made that the good performance is in part due to the fact that 16 MFCC coefficients are used, while for LPC and PSC features 5 coefficients are used. Although the dimensionality of the feature vector is an important performance parameter, the author believes that MFCCs have a clear advantage over LP-based spectral envelope modeling: Both spectral peaks and spectral valleys are equally well modeled using MFCCs, and not only spectral peaks as in the case of LP analysis. This

TABLE I. Average percent correct identification and associated 95% confidence interval (CI). Results are presented for each vowel identification method using full or diagonal covariance matrices.

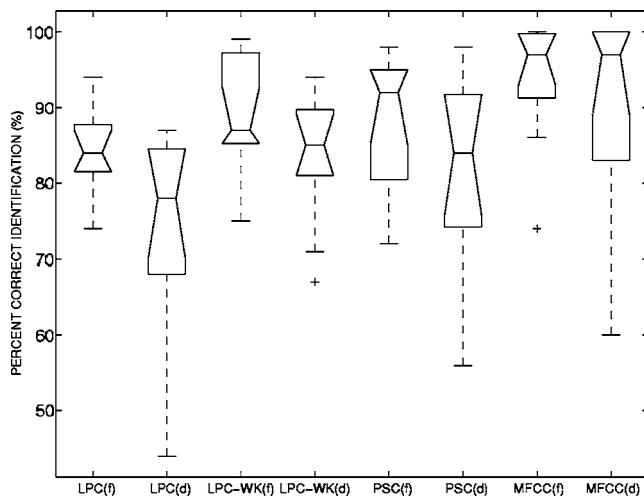|	LPC		LPC-WK		PSC		MFCC	
	Full	Diag	Full	Diag	Full	Diag	Full	Diag
Mean(%)	84.2	73.8	88.7	83.7	88.4	82.2	94.0	90.9
95% CI	[80–88]	[64–84]	[83–94]	[78–89]	[83–94]	[74–90]	[89–99]	[82–100]

FIG. 13. Boxplots denoting the identification performance of each feature set when full (*f*) or diagonal (*d*) covariance matrices are used in the Mahalanobis distance. The notched boxes represent the lower quartile, median and upper quartile values of the classification results. Whiskers extending from each box are representative of all the data and the plus sign represents outliers. The notches represent a robust estimate of the uncertainty about the medians.

fact may play an important role in the recognition performance when MFCC features are used, and suggests that valleys are perceptually relevant.

Figure 13 depicts the identification performance of the different methods as evaluated by carrying an ANOVA test on the cross-validation results. Figure 13 confirms our conclusions regarding the relative performance of LPC features and MFCC features. It also suggests that the performance of LPC-WK and PSC features are comparable, although the median resulting from PSC features is ranked higher than the median for LPC-WK features, when full covariance matrices are used in the Mahalanobis classification, in both cases.

VI. DISCUSSION

Several important conclusions can be extracted from the classification tests and experimental results. First, when the full covariance matrix is used in the Mahalanobis distance, the scores obtained for all tested feature sets compare favorably to the scores reported by other authors on vowel classification tests at high pitch frequencies. For example, using LP analysis and formant-based approaches, Palethorpe *et al.* (1996) report about 60% correct identification, and Zahorian and Jagharghi (1993) report about 72% correct identification, while the lowest scores in our tests are on the order of 84% correct identification. In static vowel recognition tests, Hillenbrand and Houde (2003) report that the best percent correct recognition accuracy has been obtained as 85.4%, 79.7%, and 77.2% for men, woman and child speakers, respectively. However, it should be noted that these authors used tests for classification of 10, 11, or even 12 vowels, a considerably more challenging task than classification of 5 vowels.

When the covariance matrix is diagonalized, the lowest performance is obtained for the plain LPC features (i.e., the first four roots of the LP polynomial, in addition to pitch

frequency), on the order of 74% correct identification, which is reasonably in line with results from other authors.

While there are no dramatic differences between the presented performance results and those from other authors, there are specific aspects that may account for certain differences. One aspect regards sampling frequency. The audio data considered in this paper is sampled at 32 kHz, while the highest sampling frequency other authors use is 20 kHz. A higher sampling frequency not only gives rise to a more extended spectral characterization, for the same pitch frequency, but also ensures a less perturbed spectral estimation due to aliasing effects. An additional aspect may be related to the fact that our testing conditions are significantly constrained, namely by limiting each utterance to a duration of exactly 100 ms corresponding to a stationary region of the natural speech recording, and also by focusing, as a consequence, on static cues rather than on a combination of static plus dynamic cues.

A second important conclusion concerns the influence of the pitch information. In fact, results present clear evidence that LP based classification can be significantly improved by interpolating between the peaks of all the partials pertaining to the harmonic structure of a voiced vowel, as was considered in the LPC-WK feature set. The main advantage is that the poles modeling the resulting spectral envelope[14] reflect better perceptual concentrations of signal energy, instead of being distracted by local spectral peaks (i.e., individual partials of the harmonic structure). This fact makes clear that by explicitly including pitch information, LP-WK features and PSC features effectively share some structural aspects. Specifically, in the case of the back vowels /a/, /o/ and /u/, and presuming male speech to simplify the discussion, the F1 and F2 formant frequencies correspond approximately to the COG and right border of the first PSC (i.e., they correspond approximately to features COG1 and PSCR) and, in the case of front vowels /e/ and /i/, F1 corresponds approximately to COG1, and F2 corresponds approximately to COG2. This relation is better illustrated using the spectral representations in Fig. 1. These conclusions also accommodate the suggestion that spectral valleys and spectral slopes on each border of each PSC have relevant perceptual information, which is a topic for further research. This hypothesis is in line with the research of Hillenbrand *et al.* (2006), which has indicated that in evaluating speech intelligibility, the fine structure of the spectrum has an advantage over a smoothed spectrum preserving the broad peaks of the spectral envelope.

An additional important conclusion is that results also provide clear evidence that MFCCs perform better than any other tested feature set in the vowel recognition tests at high pitch. The minimum performance advantage is on the order of 5% (when a full covariance matrix is used in the Mahalanobis distance) and the maximum performance advantage is on the order of 17% when the covariance matrix is diagonalized. These results are very impressive in the face of the fact that MFCCs discard pitch information. It must be noted however that contrary to plain formant estimation based on LP analysis, MFCCs are not distracted by pitch harmonics since there are two stages of spectral smoothing in the computation of MFCC coefficients. One smoothing stage occurs in the

FFT analysis since 50% zero-padding is used. Another smoothing stage occurs in the Mel-scale frequency estimation since the magnitude of FFT frequency bins are accumulated in each Mel band according to a triangular smoothing function (Davis and Mermelstein, 1980). Thus, this spectral smoothing avoiding (distracting) local spectral detail (such as harmonic peaks) is equivalent to the strategy followed in deriving the LPC-WK features, and is equivalent to the Hillenbrand *et al.* (2006) approach in computing a *harmonic envelope* in order to build a spectral envelope synthesizer. The main difference between these alternative smoothing criteria lies in the underlying spectral resolution.[15] The author believes this aspect plays a relevant role in contributing to the performance of MFCCs. As explained in Sec. V, another important aspect is the fact that MFCC coefficients model the smooth envelope of the Mel-scale spectrum using cepstral techniques which capture information reflecting both spectral peaks and spectral valleys.

VII. SUMMARY AND CONCLUSION

In this paper the concept of perceptual spectral cluster has been proposed to model the perceptual processing of the human auditory system in extracting information from stationary voiced sounds, and specifically in identifying vowel sounds at high pitch. According to this concept, an hypothesis is formulated that the human identification of a voiced vowel builds on the preliminary identification of the pitch, harmonic structure, and timbre of the vowel sound. It is further considered that the two most prominent clusters of spectral partials contain sufficient information that is captured by the HAS and allow one to discriminate high pitched vowels, in a robust and reliable way.

An algorithm has been proposed that allows the estimation of PSC boundaries in the magnitude spectrum of the vowel signal, and that allows the estimation of PSC (static) features. Using an analysis/synthesis processing environment, we performed a large number of experimental tests assessing the impact on vowel identification by a human, as a result of a spectral manipulation reflecting the influence of PSC features. The four most sensitive PSC features found (besides pitch frequency) have been selected and have been evaluated in automatic vowel identification tests using a classification procedure based on the Mahalanobis distance. In order to benchmark the test results, three alternative feature sets have been used and automatic recognition tests have been repeated using the same methodology and database. The feature sets consist of the angular values of the first four roots of the LP polynomial (LPC), the angular values of the first four roots of the LP polynomial obtained after spectral smoothing using the magnitudes of the harmonic partials (LPC-WK), and MFCC coefficients. The database contains the utterances of 5 Portuguese tonic vowels uttered by 44 speakers, 27 of whom are child speakers with a predominance of 5 and 9 year old speakers.

The automatic identification tests have revealed that the best discriminating PSC features are center of gravity and right border of the first PSC. When compared to other feature sets in terms of automatic identification performance, the selected PSC features compare favorably to LPC features, show a performance that is comparable to that of LPC-WK features, and exhibit a performance that is lower than that due to MFCC coefficients.

While the PSC concept is inspired by the actual signal processing taking place on the HAS for stationary vowel sounds, the selected PSC static features provide only about 82%–84% correct vowel recognition in test conditions (100 ms stationary sounds) where the human recognition accuracy, in practical terms, is near 100%. These results are, however, encouraging and recommend further research on alternative PSC features, or on a perceptually more appropriate utilization of the pitch information [in normalizing vowel sounds for example, which is a hypothesis supported by several authors, e.g., de Cheveigné and Kawahara (1999), p. 3506], in articulation with PSC features. Not only is this approach in line with Klatt's (1982) conclusion that "in phonetic distance judgment, a metric is needed that attends to the locations of prominent energy concentrations" but it may also prove to have substantial advantages over MFCC coefficients, namely by facilitating auditory object separation using accurate pitch estimation and harmonic analysis and segregation. In theory, this would would make it possible for PSC features to correctly recognize different simultaneous vowel sounds captured by a single microphone, which is not possible using MFCC features without pitch estimation.

[1]A description of the project and its realizations can be found at the Internet site 〈http://www.inescporto.pt/cienciaviva/〉 (last viewed on May 12th 2007).

[2]Besides this main conclusion, the authors also give an important hint: some measure of gross spectral shape might be a more appropriate representation of vowels in the auditory system than spectral peaks.

[3]The authors have also concluded that to a large extent, results support the theory of the acoustic-phonetic invariance of vowels.

[4]Front vowels are characterized by the fact that the F1 formant is much lower than the F2 formant.

[5]Back vowels are characterized by the fact that the first two formants (F1, F2) are very close to each other.

[6]This process is also accompanied by top-down perceptual mechanisms (and not only bottom-up), but the focus in this paper is placed on the integration of multiple features in this process.

[7]This aspect probably also plays a relevant role in the well-known "cocktail party effect" (Moore, 1989).

[8]The sound recordings were normalized in magnitude to 80% of the full dynamic range, prior to digital signal processing.

[9] This value has been found experimentally so as to preserve vowel identity by a human listener, after resynthesis of the vowel with its spectrum modified in magnitude according to the described algorithm.

[10] The Wiener-Khintchine theorem states that the autocorrelation and the spectral density form a Fourier pair.

[11] This tool box can be found at the following Internet site: ⟨http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/⟩ (last viewed on May 12th 2007).

[12] This database is available from the author upon request.

[13] These studies also suggest that good identifiability by humans of vowels when F0 is high, confirms the "sufficient contrast hypothesis," which states that female formant patterns are more dispersed in the vowel space and this provides a compensatory effect.

[14] Conceptually, this spectral envelope is similar to the *harmonic envelope* suggested by Hillenbrand *et al.* (2006).

[15] As pointed out by de Cheveigné and Kawahara (1999), a pertinent issue is to avoid that "smoothing replaces data that are incomplete but correct, with data that are complete but incorrect."

Bladon, R. W. (**1982**). "Arguments against formants in the auditory representation of speech," in *The Representation of Speech in the Peripheral Auditory System*, edited by R. Carlson and B. Granstrom (Elsevier Biomedical, Amsterdam), pp. 95–102.

Bruce, I. C., Karkhanis, N. V., Young, E. D., and Sachs, M. B. (**2002**). "Robust formant tracking in noise," in IEEE International Conference on Acoustics, Speech and Signal Processing, pp. I281–I284, Orlando, Florida.

Chen, B., and Loisou, P. C. (**2004**). "Formant frequency estimation in noise," in IEEE International Conference on Acoustics, Speech and Signal Processing, pp. I581–I584, Montreal, Canada.

Chistovich, L., and Lublinskaja, V. (**1979**). "The center of gravity effect in vowel spectra and critical distance between the formants: Psychoacoustical study of perception of vowel-like stimuli," Hear. Res. **1**, 185–195.

Davis, S. B., and Mermelstein, P. (**1980**). "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Trans. Acoust., Speech, Signal Process. **28**, 357–366.

de Cheveigné, A., and Kawahara, H. (**1999**). "Missing-data model of vowel identification," J. Acoust. Soc. Am. **105**, 3497–3508.

Diehl, R. L., Lindblom, B., Hoemeke, K. A., and Fahey, R. P. (**1996**). "On explaining certain male-female differences in the phonetic realization of vowel categories," J. Phonetics **24**, 187–208.

Dusan, S., and Rabiner, L. (**2005**). "On integrating insights from human speech perception into automatic speech recognition," in the Ninth European Conference on Speech Communication and Technology (Interspeech-2005), Lisbon, Portugal, pp. 1233–1236.

Fant, G. (**1970**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Ferreira, A., and Sinha, D. (**2005**). "Accurate and robust frequency estimation in the odft domain," in 2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, pp. 203–206.

Ferreira, A. J. S. (**1996a**). "Audio spectral coder," 100th Convention of the Audio Engineering Society, Copenhagen, Denmark.

Ferreira, A. J. S. (**1996b**). "Perceptual coding of harmonic signals," 100th Convention of the Audio Engineering Society, Copenhagen, Denmark.

Ferreira, A. J. S. (**1998**). "Spectral coding and post-processing of high quality audio," Ph.D. thesis, Faculdade de Engenharia da Universidade do Porto-Portugal, Porto, Portugal, http://telecom.inescn.pt/doc/phd_en.html (last viewed on May 12th 2007).

Ferreira, A. J. S. (**2001**). "Accurate estimation in the odft domain of the frequency, phase and magnitude of stationary sinusoids," in 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, pp. 47–50.

Ferreira, A. J. S. (**2005**). "New signal features for robust identification of isolated vowels," in the Ninth European Conference on Speech Communication and Technology (Interspeech-2005), Lisbon, Portugal, pp. 345–348.

Flexer, A. (**2006**). "Statistical evaluation of music information retrieval experiments," J. New Music Res. **35**, 113–120.

Goldstein, L. G. (**1973**). "An optimum processor theory for the central formation of the pitch of complex tones," J. Acoust. Soc. Am. **54**(6), pp. 1496–1516.

Hermansky, H. (**1990**). "Perceptual linear predictive (plp) analysis of speech," J. Acoust. Soc. Am. **87**, 1738–1752.

Hermansky, H., Hanson, B., and Wakita, H. (**1985**). "Low-dimensional representation of vowels based on all-pole modeling in the psychophysical domain," Speech Commun. **4**, 181–187.

Hess, W. (**1983**). *Pitch Determination of Speech Signals—Algorithms and Devices* (Springer, Berlin).

Hillenbrand, J. M., and Houde, R. A. (**2003**). "A narrow band pattern-matching model of vowel perception," J. Acoust. Soc. Am. **113**, 1044–1055.

Hillenbrand, J. M., Houde, R. A., and Gayvert, R. T. (**2006**). "Speech perception based on spectral peaks versus spectral shape," J. Acoust. Soc. Am. **119**, 4041–4054.

Honig, F., Stemmer, G., Hacker, C., and Brugnara, F. (**2005**). "Revisiting perceptual linear prediction (plp)," in the Ninth European Conference on Speech Communication and Technology (Interspeech-2005), Lisbon, Portugal, pp. 2997–3000.

Klatt, D. H. (**1982**). "Prediction of perceived phonetic distance from critical-band spectra—a first step," in IEEE International Conference on Acoustics, Speech and Signal Processing, Paris, France, pp. 1278–1281.

Lass, N. J., ed. (**1996**). *Principles of Experimental Phonetics* (Mosby Year Book, St. Louis, MO).

Maurer, D., Cook, N., Landis, T., and d'Heureuse, C. (**1992**). "Are measured differences between the formants of men, women and children due to f0 differences?," J. Int. Phonetic Assoc. **21**, 66–79.

Mollis, M. R. (**2005**). "Evaluating models of vowel perception," J. Acoust. Soc. Am. **118**, 1062–1071.

Moore, B. C. J. (**1989**). *An Introduction to the Psychology of Hearing* (Academic, New York).

Palethorpe, S., Wales, R., Clark, J. E., and Senserrick, T. (**1996**). "Vowel classification in children," J. Acoust. Soc. Am. **100**, 3843–3851.

Plomb, R. (**2002**). *The Intelligent Ear—On the Nature of Sound Perception* (Erlbaum, Mahwah, NJ).

Rabiner, L., and Juang, B.-H. (**1993**). *Fundamentals of Speech Recognition* (Prentice-Hall, Englewood Cliffs, NJ).

Ryalls, J. H., and Lieberman, P. (**1982**). "Fundamental frequency and vowel perception," J. Acoust. Soc. Am. **72**, 1631–1634.

Sroka, J. J., and Braida, L. D. (**2005**). "Human and machine consonant recognition," Speech Commun. **45**, 401–423.

Thorpe, C., and Watson, C. (**2000**). "Vowel identification in singing at high pitch," in Proceedings of the Eighth Australian International Conference on Speech Science and Technology, Canberra, Australia, pp. 280–286.

Vaidyanathan, P. P. (**1993**). *Multirate Systems and Filter Banks* (Prentice-Hall, Englewood Cliffs, NJ).

Wet, F., Weber, K., Boves, L., Cranen, B., Bengio, S., and Bourlard, H. (**2004**). "Evaluation of formant-like features on an automatic vowel classification task," J. Acoust. Soc. Am. **116**, 1781–1791.

Zahorian, S. A., and Jagharghi, A. J. (**1993**). "Spectral-shape features versus formants as acoustic correlates for vowels," J. Acoust. Soc. Am. **94**, 1966–1982.

Zwicker, E. (**1961**). "Subdivision of the audible frequency range into critical bands," J. Acoust. Soc. Am. **33**, 248–284.

Compensatory responses to loudness-shifted voice feedback during production of Mandarin speech

Hanjun Liu and Qianru Zhang
Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208

Yi Xu
Department of Phonetics and Linguistics, University College London, London, United Kingdom

Charles R. Larson[a)]
Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208

Previous studies have demonstrated that perturbations in voice pitch or loudness feedback lead to compensatory changes in voice F_0 or amplitude during production of sustained vowels. Responses to pitch-shifted auditory feedback have also been observed during English and Mandarin speech. The present study investigated whether Mandarin speakers would respond to amplitude-shifted feedback during meaningful speech production. Native speakers of Mandarin produced two-syllable utterances with focus on the first syllable, the second syllable, or none of the syllables, as prompted by corresponding questions. Their acoustic speech signal was fed back to them with loudness shifted by ±3 dB for 200 ms durations. The responses to the feedback perturbations had mean latencies of approximately 142 ms and magnitudes of approximately 0.86 dB. Response magnitudes were greater and latencies were longer when emphasis was placed on the first syllable than when there was no emphasis. Since amplitude is not known for being highly effective in encoding linguistic contrasts, the fact that subjects reacted to amplitude perturbation just as fast as they reacted to F_0 perturbations in previous studies provides clear evidence that a highly automatic feedback mechanism is active in controlling both F_0 and amplitude of speech production. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2773955]

PACS number(s): 43.72.Dv, 43.70.Mn, 43.70.Jt, 43.70.Gr [AL]   Pages: 2405–2412

I. INTRODUCTION

There have been several studies of the mechanisms that control voice intensity. Considering peripheral mechanisms, subglottal air pressure, air flow, glottal impedance, voice fundamental frequency (F_0) and vocal tract impedance all interact to affect vocal intensity (Isshiki, 1964; Koyama *et al.*, 1969; Titze and Sundberg, 1992). The variations in vocal intensity that occur during speech are the result of interactions between the pressure-relaxation forces of the respiratory system, and respiratory and laryngeal muscle contractions (Draper *et al.*, 1959; Hirano and Ohala, 1969; Ladefoged and Loeb, 2002). Neural mechanisms of voice intensity control are less well understood. Lombard was the first to demonstrate the importance of auditory feedback on the control of intensity (see Lane and Tranel, 1971). It was found that the presence of environmental noise affected voice intensity where speakers raised their vocal intensity to make themselves heard over the noise level. Similarly, the phenomenon of side-tone amplification demonstrated that if a speaker's voice is amplified above a normal level, the speaker will reduce his or her intensity; if the feedback level

of a person's voice is reduced, speakers will raise their intensity (Chang-Yit *et al.*, 1975; Garber *et al.*, 1976; Lane and Tranel, 1971; Siegel and Pick Jr., 1974).

As important as these studies are, the research paradigm used in them raises questions on their interpretation. In the typical paradigm, a person is instructed to read a passage in the presence of noise, side-tone amplification, or both side-tone amplification and noise. Generally these auditory feedback variables are present for the duration of the speaking task, and the intensity adjustments made by the speaker are considered to be automatic. In this sense, the word automatic does not necessarily mean reflexive, but rather an adjustment a speaker would naturally make to increase the effectiveness in communicating with others. Because of the reliability of the side-tone amplification effect, it was suggested that it may reflect a "fundamental characteristic of speech regulation" (Chang-Yit *et al.*, 1975, p. 324) to maximize the communicative effectiveness by increasing the signal to noise ratio. Despite the importance of regulating one's voice in an attempt to overcome noise or distance, it is not clear from the above studies whether subjects monitor voice feedback and make corrections online in case the production does not match that which was intended.

Two recent studies have demonstrated that auditory feedback seems to play a role in the online control of voice intensity during vowel production. Heinks-Maldonado and

_____

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

Houde (2005) and Bauer et al. (2006) demonstrated that during vowel productions, speakers would respond to brief (e.g., 200 ms) perturbations in voice loudness feedback by making changes in their voice amplitude. The latencies of the responses, with a mean value between 150 and 300 ms, coupled with the fact that the direction of the responses is generally opposite to the stimulus direction, regardless of whether the stimulus is an increase or decrease in loudness feedback, suggests these responses are reflexive in nature. These studies are similar to those that have been conducted to study the effects of voice pitch feedback perturbations on voice F_0 control, which are also thought to be reflexive in nature (Bauer and Larson, 2003; Burnett et al., 1998; Hain et al., 2000). Moreover, the fact that responses to pitch perturbations are observed in speech (Donath et al., 2002; Jones and Munhall, 2002; Natke et al., 2003), and increase in magnitude in some speaking conditions (Chen et al., 2007; Xu et al., 2004), raises the possibility that voice amplitude responses to perturbations in loudness feedback may also be present in speech. A third line of research demonstrating the importance of auditory feedback in speech comes from studies showing long term adaptation to changes in formant frequencies (Houde and Jordan, 1998; Purcell and Munhall, 2006).

However, the role of amplitude in speech is unclear, and whether or not amplitude is controlled like F_0 may depend on the importance of amplitude control during speech. While there is evidence that overall amplitude is actively controlled by speakers and that auditory feedback plays a role in this process (Bond and Moore, 1994; Chang-Yit et al., 1975; Dreher and O'Neill, 1958; Garber et al., 1976; Lane et al., 1995; Lane et al., 1997; Leder et al., 1987; Siegel and Kennard, 1984; Svirsky et al., 1992; Van Summers et al., 1988), there have also been findings suggesting that, unlike F_0, amplitude is not highly effective in conveying communicative functions such as lexical stress, or focus (Fry, 1958; Turk and Sawusch, 1996). A recent analysis of natural speech databases, however, has found amplitude to be correlated with the perception of prominence (Kochanski et al., 2005). Regardless of its actual function, it is an open question as to whether amplitude is rapidly adjusted in response to perturbations in voice loudness during the production of meaningful speech. If voice amplitude is adjusted online in response to perturbations in auditory feedback, it would suggest that amplitude control during speech is important for the expression of linguistic functions such as focus or word stress. The purpose of the present study was to investigate whether speakers compensate for perturbations in voice loudness feedback during speech.

The Mandarin speech stimuli that were investigated were designed to differ in their focus patterns. Focus is discourse-motivated emphasis, and is known to be accompanied by expansion of pitch range in the focused word and suppression of pitch range in the postfocus words (Xu, 1999; Xu and Xu, 2005). The pitch range expansion in a final focus is much smaller than in an earlier focus, which makes its percept less salient (Liu and Xu, 2005). Although there have not been systematic data, the pitch range changes due to focus may also be accompanied by amplitude variations. It is also possible that such amplitude variations are actively controlled. If they are actively controlled, and if auditory feedback plays a role in this control process, then responses to loudness-shifted voice feedback during a focused syllable should be greater than if the syllable is not focused.

II. METHODS

A. Subjects

Ten native speakers of Mandarin (six males and four females; ages 19–30), most of whom were students at Northwestern University, served as subjects. All subjects reported normal hearing, and none reported a history of neurological or communication disorders. All signed informed consent approved by the Northwestern Institutional Review Board and were paid for their participation.

B. Apparatus

Subjects were seated in a sound-treated room and wore Sennheiser headphones with attached microphone (model HMD 280) throughout the testing. They were asked to speak aloud the experimental stimuli at approximately 70 dB sound pressure level (SPL), self-monitoring their voice loudness from a Dorrough Loudness Monitor (model 40-A) placed 0.5 m in front of them. This monitor provided the subjects with visual feedback on their voice amplitude and helped them to maintain a relatively constant level throughout the testing. The rapid changes in the visual display, which coincided with changes in voice amplitude, were too fast for the subjects to respond to and thus do not affect the results. This feedback merely helped the subjects to maintain a relatively constant amplitude level throughout the testing. The vocal signal from the microphone was amplified with a Mackie mixer (model 1202), processed for loudness shifting with an Eventide Eclipse Harmonizer, mixed with 40 dB SPL pink masking noise with a Mackie mixer (model 1202-VZL), further amplified with Crown D75 amplifier and HP 350 dB attenuators at 80 dB SPL, and sent back to the headphones. The harmonizer was controlled with MIDI software (Max/MSP v.4.1 by Cycling 74) from a laboratory computer. Acoustic calibrations were made with a Brüel & Kjær sound level meter (model 2250) and in-ear microphones (model 4100). There was a gain of 10 dB SPL between the subject's voice amplitude, measured 2.5 cm from the mouth, and the feedback loudness measured at the input to the ear canal. The voice output signal, feedback and control pulses (TTL) were digitized at 10 kHz, low-pass filtered at 5 kHz and recorded on a laboratory computer utilizing Chart software (AD Instruments). Data were analyzed using event-related averaging techniques in Igor Pro (Wavemetrics, Inc., Lake Oswego, OR).

C. Procedures

The disyllabic sequence produced by the subjects was /ba1 ma1/, meaning "the eighth aunt." This phrase was produced in response to three different questions the subject would hear over the headphones, as shown in Table I. Each of the three questions required the subjects to produce the

TABLE I. Disyllabic list of questions and responses, where numerals 1, 2, 4 represent the High, Rising, and Falling tones, and the underscored syllable is focused.

Question	Response
/shui?/ "Who?"	/ba1 ma1/ (No Focus)
/ba4 ma1?/ "Father and mother?"	/<u>ba1</u> ma1/ (1st Focus)
/ba1 yi2/? "The eighth aunt in mother's family?"	/ba1 <u>ma1</u>/ (2nd Focus)

phrase with one of three different focus patterns: first syllable (1st Focus), second syllable (2nd Focus) or neither syllable (No Focus) (Xu, 1999).

Each subject completed a total of 180 trials that were divided into three groups of 60 trials with intervening rest periods. In each group of trials, 60 questions requiring 60 responses were presented. The three questions were randomly distributed among the group of 60. During the production of a group of 60 trials, voice loudness feedback was either increased, decreased or not changed (control trials). These stimuli were presented randomly so that the subjects could not predict which type of stimulus, or control (no stimulus), would occur on each trial. Thus, a subject received 20 increases in loudness feedback, 20 decreases and 20 control trials for each of the three questions. The stimuli were ±3 dB SPL perturbations (200 ms duration) in voice loudness feedback beginning 160 ms after vocal onset. Figure 1 shows example voice amplitude contours during production of the three phrases (solid lines) superimposed on the perturbed loudness feedback signal (dashed lines). In the first example there was a decrease in feedback loudness, in the middle example there was no change in feedback loudness (control) and in the 3rd example there was an increase in feedback loudness. During the recording, three TTL pulses were recorded that indicated the type of loudness shift stimulus (upward, downward, or control) that was presented. During data acquisition, the experimenter listened to the recorded vocal responses to insure that the subject responded with the correct response to each question. Incorrect responses were discarded from further analysis, and extra trials were run to insure that 180 total trials with the correct responses were obtained for each subject.

D. Data analysis

Digitized signals were analyzed by converting the voice signal to a root-mean-square (rms) voltage signal calculated using a 50 ms sliding window
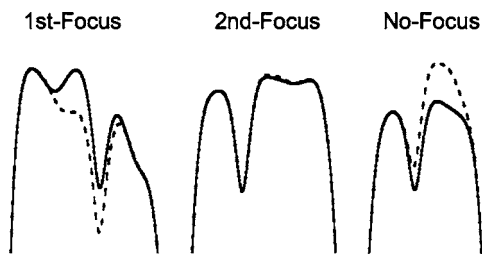

FIG. 1. Voice amplitude contours during production of the three phrases, 1st Focus, 2nd Focus and No Focus. Solid lines represent voice amplitude and dashed lines represent voice loudness feedback. These examples are for single productions. Time and magnitude scales are in relative units.

$$\text{rms}(x) = \sqrt{\frac{1}{N}\sum_{n-25}^{n+25} x^2},\qquad(1)$$

where x is the value of each data point, and N is the total number of data points. Voice rms voltage measures were then converted to dB SPL to reflect actual voice amplitude using the following formula:

$$\text{Voice(dB)} = 20 \times \log(\text{rms}(x)/c) + 70,\qquad(2)$$

where c equals 0.228, which is the rms voltage corresponding to a vocal level of 70 dB SPL that was obtained through calibration procedures. This wave and the TTL pulses were displayed on a computer screen. An operator marked the onset and offset of each vocalization (phrase) according to the voice wave form signal. The durations of all vocal signals along with the accompanying TTL signal representing timing and direction of the stimulus were then time normalized. Time normalization (linear interpolation) was done to reduce temporal variations in the speech signals and thereby reduce variability in the subsequent averaged trials. For each subject, an ensemble average of each set of test and control trials was generated for each speech and stimulus direction condition by triggering the averaging program at the onset of each of the TTL pulses.

A point-by-point series of t tests were then conducted between all the test and respective control trials for a given condition (Chen et $al.$, 2007; Xu et $al.$, 2004). The result of the t test was a set of "p" values representing the significance of the t test for each time point of the set of control and test waves. Wherever the "p wave" decreased below a value of 0.02 following the stimulus onset with a delay of at least 60 ms and remained low for at least 50 ms, the crossing point was designated as the onset of a response (latency). The temporal constraints imposed by these criteria guarded against spurious significant differences (Chen et $al.$, 2007). The point where the p wave increased to a value greater than 0.02 was defined as the response termination. A "difference" wave was then calculated by subtracting the average control wave from the averaged increasing and decreasing loudness stimulus test waves for each subject and each condition. Thus for each focus and perturbation condition there was a "difference wave" representing the response to an upward shift in loudness feedback compared with the control condition, and one representing the response to a downward shift in loudness feedback compared with the control condition. Using the times noted in the analysis of the p wave crossing the value of 0.02, a program then measured the peak (or trough) magnitude of the difference wave. If responses failed to reach significance within the above-described temporal parameters, they were designated as nonresponses. The time stamps of the p wave crossings along with the magnitude measures of the difference waves were tested for significance with a repeated-measures analysis of variance (ANOVA) (SPSS, v. 11.0). For statistical analysis, nonresponses were replaced by the mean value calculated from the measured data from other subjects for that condition. This procedure allowed us to use a repeated-measures design. Assumptions

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Liu et $al.$: Voice amplitude control in Mandarin   2407

TABLE II. Total number of "following" (FOL), "opposing" (OPP), and "nonresponse" (NR) across two stimulus directions.

	Up	Down	Total
FOL	11	9	20
OPP	14	16	30
NR	5	5	10
Total	30	30	60

of a normal distribution, homogeneity of variance, compound symmetry and circularity for a repeated measures ANOVA were met.

III. RESULTS

From ten subjects across the three experimental conditions and two stimulus directions (10*3*2), there were 60 possible responses. Tables II and III list the numbers of opposing, "following" and nonresponses across the two stimulus directions and three intonation patterns. Approximately 33.3% of the responses "followed" the stimulus direction. About 16.7% of the responses did not meet our criteria of validity and were declared to be nonresponses. 50% of responses opposed the direction of the loudness-shift stimulus. The percentage of valid responses did not vary greatly across the three experimental conditions. Also the numbers of valid responses by stimulus direction did not differ greatly (14 Up and 16 Down).

Figure 2 shows examples of the average voice amplitude contours for each of the three questions and stimulus direction. In each plot, the responses to loudness shift stimuli (thick lines) are shown along with the average control curves (thin lines). At the bottom of the plots, the square brackets represent the timing and direction of the stimulus. In all the plots, the first syllable is clearly separated from the second syllable by a drop in SPL. Due to the differences in the emphasis patterns, the duration of syllables /ba/ and /ma/ changed. The first syllable is longer when emphasis is on the first syllable than when it is on the second syllable or when there is no emphasis.

In Fig. 2 the drop in SPL associated with the transition to the second syllable occurred right after the stimulus onset. The stimuli persist into the second syllable, and the response, indicated by the divergence between the test and control contours, can be easily seen during the second syllable. The amount of the separation reflects the response magnitude. Arrows indicate the time at which the response magnitude was measured. Responses to increasing loudness stimuli are shown on the left and decreasing loudness stimuli on the right. In Fig. 2, the response to the increasing stimulus in the

TABLE III. Total number of "following" (FOL), "opposing" (OPP), and "nonresponse" (NR) across three phrase types.

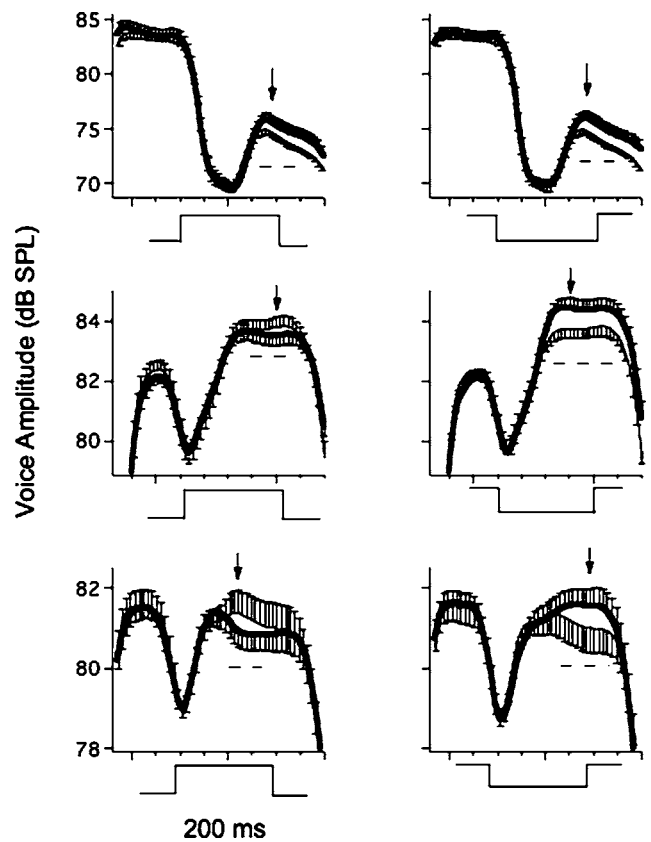|	1st Focus	2nd Focus	No Focus	Total
FOL	6	7	7	20
OPP	9	11	10	30
NR	5	2	3	10
Total	20	20	20	60

FIG. 2. Control (thin black line) and test average waves (thick black line) during 1st- Focus, 2nd-Focus, and No-Focus sequences (from top to bottom). Contours cover the entire duration of the utterances. The vertical arrow indicates time where the response magnitude was measured. Error bars represent the standard error of the mean for a single direction. Curves at the bottom indicate the time and the direction of the stimulus. Horizontal dashed lines indicate time period where the two waves differed significantly.

1st-Focus condition "follows" the direction of the stimulus. All other responses in this figure oppose the direction of the stimulus and are compensatory in nature.

Figures 3 and 4 show box plots of the response magnitude and latency across three phrase types and two stimulus directions, respectively. Values of response magnitudes and latencies are shown in Tables IV and V. Two-factor repeated-measures ANOVAs were performed on measures of response magnitude and latency across the factors phrase type and stimulus direction. For statistical analysis, measures for both the compensating and following responses were grouped together because there were only 50 total responses, and there was a relatively large number of following responses compared to previous studies (Chen et al., 2007; Xu et al., 2004). Moreover, there were no statistically significant differences in the measures of response magnitude ($F(1,49)=1.743$, $p=0.191$) or latency ($F(1,49)=0.964$, $p=0.330$) between the groups. For the response magnitude, a significant main effect was found for phrase type ($F(2,18)=3.923$, $p=0.039$, Fig. 3 and Table IV) but not for stimulus direction ($F(1,9)=0.991$, $p=0.345$; Fig. 4 and Table V). A post hoc test indicated that the responses in the 1st-Focus phrase were significantly larger than those in the 2nd-Focus phrase ($p=0.032$; post hoc Bonferroni; Table IV). Response magnitudes for the 2nd Fo-
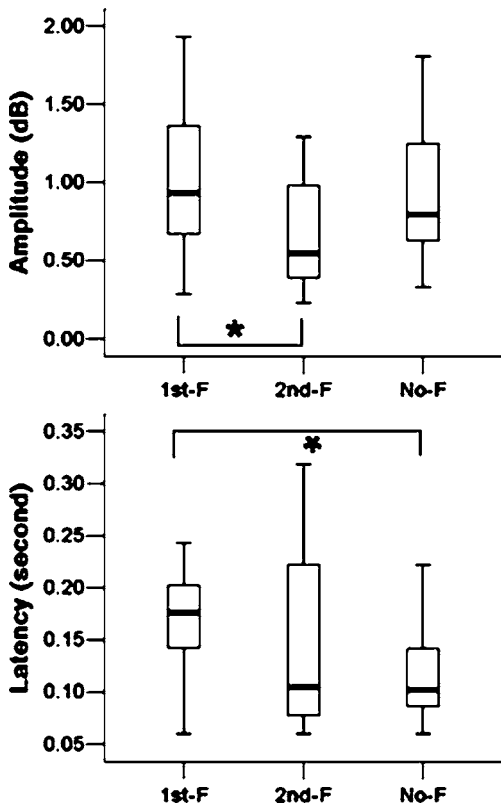
FIG. 3. Box plots illustrating the response magnitude (the top row) and the latencies (the bottom row) as a function of phrase type. Box definitions: middle line is median, top and bottom of boxes are 75th and 25th percentiles, whiskers extend to limits of main body of data defined as high hinge +1.5 (high hinge−low hinge), and low hinge −1.5 (high hinge−low hinge) (Data Desk; Data Description). Asterisk bracket indicates significance between measures.
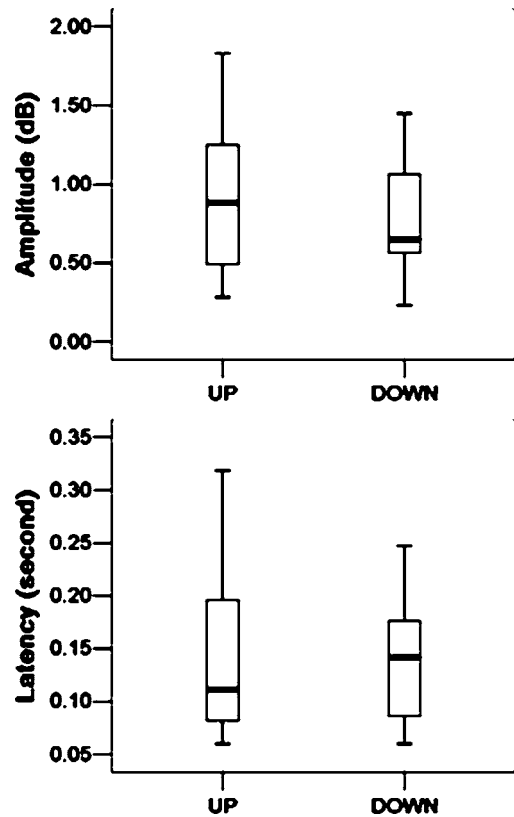

FIG. 4. Box plots illustrating the response magnitude (the top row) and the latencies (the bottom row) as a function of stimulus direction.

cus and No Focus were not significantly different. No significant interaction was found between phrase type and stimulus direction.

For response latency, two-way repeated-measures ANOVAs revealed significant main effects of phrase type $(F(2,18)=4.824, p=0.021)$. Post hoc Bonferroni tests indicated that the latencies for the 1st-Focus phrase were significantly longer than those for the No-Focus phrase $(p=0.039)$ (see Fig. 3 and Table V). Latencies for the 2nd-Focus and No-Focus conditions were not statistically different. Response latencies did not differ significantly as a function of stimulus direction.

IV. DISCUSSION

It is known that amplitude variation in speech is highly dependent on the characteristics of the speech sounds (Fant, 1960; Stevens, 1998) as well as lung volume (Ladefoged and Loeb, 2002). For example, other things being equal, ampli-

tude is negatively related to the height of the vowel (hence positively related to F1)—the higher the vowel, the lower the amplitude. Also, amplitude is closely related to speech sound type: it is much higher in vowels than in consonants. These facts make it difficult for amplitude to be a direct object of control in speech production, as is evident in the findings that amplitude is not highly effective in encoding linguistic contrasts (Fry, 1958; Turk and Sawusch, 1996). Thus the question as to whether human subjects would exhibit compensatory responses to amplitude perturbation similar to those to pitch perturbation is highly interesting. The similarity in responses to loudness perturbation found in the present study and responses to pitch perturbations found in previous studies (Chen et al., 2007; Xu et al., 2004) suggest the existence of a highly automatic feedback mechanism that assists in the control of various aspects of the vocal output, including both F_0 and amplitude of voice.

The present study was designed to investigate vocal responses to loudness-shifted feedback during the production of Mandarin speech. In this paradigm, subjects varied their focus across three different phrases. At the same time, their voice loudness feedback was experimentally manipulated for

TABLE IV. Average response magnitudes (SD) across three phrase types and two stimulus directions.

Phrase		1st Focus	2nd Focus	No Focus
Direction	Up	1.07 (0.50)	0.66 (0.34)	0.99 (0.51)
	Down	1.02 (0.54)	0.67 (0.37)	0.77 (0.24)

TABLE V. Average response latencies (SD) across three phrase types and two stimulus directions.

Phrase		1st Focus	2nd Focus	No Focus
Direction	Up	171 (56)	168 (106)	100 (49)
	Down	159 (54)	130 (70)	135 (55)

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Liu et al.: Voice amplitude control in Mandarin   2409

TABLE VI. Average amplitude (SD) of the two syllables across three phrase types.

Phrase		1st Focus	2nd Focus	No Focus
Syllable	Syllable 1	84.62 (4.2)	83.85 (3.38)	82.27 (3.11)
	Syllable 2	81.16 (4.10)	85.13 (3.94)	82.26 (3.17)

short durations (200 ms). The 3 dB perturbations were added to or subtracted from the amplitude of the subjects' productions and were clearly distinguishable from the nonperturbed, control trials (see Fig. 1). The onset of the perturbation was presented usually during the first syllable, but for some subjects who increased their rate of speech, the perturbation onset occurred during the transition between the first and second syllables. In all cases the perturbation ended during the second syllable, and the response also occurred during the second syllable.

Results showed that, similar to pitch-perturbation studies, the subjects responded to increasing and decreasing loudness perturbations by changing their voice amplitude in response to the stimuli. Half the responses were compensatory, i.e., they were opposite in direction to the stimulus, and close to one third of the responses "followed" the direction of the stimulus. It is unknown why the number of "following" responses was greater than in previous studies of voice pitch or loudness feedback (Chen et al., 2007; Xu et al., 2004), particularly since it is not known what causes such responses in the first place. In a previous paper on responses to loudness perturbations during vowel phonations, greater numbers of "following" responses were speculated to result from the fact that the stimuli (−3 dB, 200 ms duration) were difficult to perceive (Larson et al., 2007). This speculation is supported by the report that 100% of opposing responses were found in the case of 10 dB stimuli during vowel productions (Heinks-Maldonado and Houde, 2005). Also the fact that 33% of the responses were "following" in the present study while only 19% (Sivasankar et al., 2005) and 20% (Larson et al., 2007) were "following" for sustained vowels indicates that it may be more difficult to perceive the direction of short duration loudness perturbations during speech production compared to vowel phonations. Another possible explanation was that the subjects may have used the feedback signal as their choice of referent when making comparisons with their intended voice amplitude production. Using the feedback signal as the referent, as in matching a piano note while singing would cause a "following" response (Hain et al., 2000). Each of these explanations may also apply to the results of the present study. The important, but still unanswered question regarding these responses is why subjects produce them under certain conditions and not others.

The magnitudes of the compensatory responses to perturbed loudness feedback during speech imply that the response magnitude is dependent on the relative voice amplitude at the moment of the stimulus. This is seen in the finding that responses for the 1st-Focus phrase were significantly larger than those for the 2nd-Focus phrase. As can be seen in Table VI, in the 1st-Focus phrases the mean ampli-

tude of the first syllable was 3.46 dB higher than that of the second syllable, during which the response peaked in amplitude. In the 2nd-Focus phrases, the mean amplitude of the first syllable was 1.28 dB *lower* than that of the second syllable. Thus it is possible that the larger amplitude of the second syllable in the 2nd-Focus phrases may have partially inhibited the compensatory mechanism, as the underlying amplitude is rapidly rising during the response to the perturbation. It is also possible that, because the first syllable is shortened when focus is on the second syllable, as can be seen in Fig. 2, there is greater overlap of the stimuli with the second syllable than in the 1st-Focus condition. This would have increased the likelihood that the response peaked while the underlying amplitude was still increasing, thus affecting the accuracy of the response measurement. Regardless, the evidence is sufficient that in Mandarin, auditory feedback in the form of both pitch and loudness feedback is used both during the production of lexical tones (Xu et al., 2004) and focus in speech.

Comparison of results between this and previous studies on pitch- and loudness-shifted voice feedback suggests similarities in the mechanisms underlying the responses. Specifically, response magnitudes to pitch- or loudness-shifted feedback were less than the stimulus magnitudes. Although direct comparisons between responses to pitch-shifted and loudness-shifted feedback are not possible because of their different acoustical dimensions, a rough comparison can be made if response magnitudes are treated as a percent of the stimulus (hereafter "% response magnitude"). In several pitch-shift studies, % response magnitudes for a 100-cent stimulus varied from 10% to 30% (Bauer and Larson, 2003; Burnett et al., 1998; Hain et al., 2000). With a pitch-shift stimulus of 25 cents, % response magnitudes approached 100% (Larson et al., 2001). For the two previous studies of loudness-shifted feedback, % response magnitudes varied from 0.06% (0.61 dB for a ±10 dB stimulus) (Heinks-Maldonado and Houde, 2005) to about 90% (0.9 dB for a ±1 dB stimulus) (Bauer et al., 2006). In the present study, the 2nd-Focus condition yielded the lowest mean response magnitude of 0.66 dB or 22% response magnitude, while the 1st-Focus condition yielded the highest, 1.07 dB SPL or 36% response magnitude. In both pitch- and loudness-shift studies, the largest % response magnitudes occurred with the smallest stimuli. As stimulus magnitude increased, % response magnitude decreased. This general finding suggests that responses to perturbed auditory are optimally suited to correct for small variations in voice pitch or loudness feedback.

Another similarity between this and previous studies is the fact that response magnitudes and latencies varied as a function of the vocal task. Natke et al. (2003) demonstrated larger responses to pitch-shifted feedback during singing compared to speech. Xu et al. (2004) demonstrated larger responses and shorter latencies to pitch-shifted feedback when the stimulus was presented prior to a major change in the tone, e.g., the transition from a high to a falling tone. Chen et al. (2007) showed that in English speech, larger and quicker responses to pitch-shifted feedback occurred when downward pitch perturbations were presented prior to a rise

in voice F_0. In the present study, responses to the loudness-shifted feedback were significantly larger for the 1st-Focus phrase than those for the 2nd-Focus phrase. In addition, latencies were significantly longer in the 1st-Focus than in the No-Focus pattern. Thus, the modification of response magnitudes and latencies in both pitch- and loudness-shifted feedback studies during speech, indicates the nervous system is capable of modulating the influence of auditory feedback for the control of the voice.

However, even though the present study demonstrated task-dependent modulation of voice amplitude responses to loudness-shifted feedback, the nature of these modulations is quite different than those reported in earlier pitch-shift studies. In previous pitch-shift studies, latencies generally became shorter as response magnitude increased (Chen *et al.,* 2007; Xu *et al.,* 2004). In contrast, in the present study, the larger responses in the 1st-Focus condition also had longer latencies. Similarly, the shorter latency responses in the No-Focus condition did not have larger magnitudes. Thus, with pitch-shifted feedback during speech, as responses increased in magnitude, they also became quicker. With loudness-shifted feedback, however, as responses became larger, they became slower. These differences between simultaneous changes in response magnitude and latency may suggest fundamental differences in the way the nervous system uses voice loudness feedback for control of voice amplitude vs. pitch feedback for control of voice F_0 during speech. Further research should be addressed to this difference.

V. CONCLUSION

Results of the present study showed that native speakers of Mandarin made compensatory responses to loudness perturbations in a manner similar to that of previously reported responses to pitch-shifted feedback, with a time delay of about 142 ms, and a magnitude about 29% of the stimulus magnitude. The finding is highly significant given that the role of amplitude variation in speech is known to be not nearly as effective as that of pitch variation (Fry, 1958; Turk and Sawusch, 1996). Compensatory responses to perturbed auditory feedback thus seem to be part of a mechanism that reacts quite automatically to any discrepancy between the anticipated and actual feedback. Results also showed that the response magnitude was smaller if the perturbation occurred when the linguistic focus was utterance final than when it was nonfinal. This provides further evidence that auditory-feedback control of vocalization is dependent on the nature of the speech signal at the time of the perturbation (Chen *et al.,* 2007; Xu *et al.,* 2004). Future studies will help to define which specific speech segments are more or less sensitive to auditory feedback.

ACKNOWLEDGMENTS

Bauer, J. J., and Larson, C. R. (**2003**). "Audio-vocal responses to repetitive pitch-shift stimulation during a sustained vocalization: Improvements in methodology for the pitch-shifting technique," J. Acoust. Soc. Am. **114**, 1048–1054.

Bauer, J. J., Mittal, J., Larson, C. R., and Hain, T. C. (**2006**). "Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude," J. Acoust. Soc. Am. **119**, 2363–2371.

Bond, Z. S., and Moore, T. J. (**1994**). "A note on the acoustic-phonetic characteristics of inadvertently clear speech," Speech Commun. **14**, 325–337.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice F0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**, 3153–3161.

Chang-Yit, R., Pick, H. L., and Siegel, G. M. (**1975**). "Reliability of sidetone amplification effect in vocal intensity," J. Commun. Dis. **8**, 317–324.

Chen, S. H., Liu, H., Xu, Y., and Larson, C. R. (**2007**). "Voice F0 responses to pitch-shifted voice feedback during english speech," J. Acoust. Soc. Am. **121**, 1157–1163.

Donath, T. M., Natke, U., and Kalveram, K. T. (**2002**). "Effects of frequency-shifted auditory feedback on voice F0 contours in syllables," J. Acoust. Soc. Am. **111**, 357–366.

Draper, M. H., Ladefoged, P., and Whitteridge, D. (**1959**). "Respiratory muscles in speech," J. Speech Hear. Res. **2**, 16–27.

Dreher, J. J., and O'Neill, J. J. (**1958**). "Effects of ambient noise on speaker intelligibility of words and phrases," Laryngoscope **68**, 539–548.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Fry, D. B. (**1958**). "Experiments in the perception of stress," Lang Speech **1**, 126–152.

Garber, S. F., Siegel, G. M., and Pick, H. L. (**1976**). "The influence of selected masking noises on Lombard and sidetone amplification effects," J. Speech Hear. Res. **19**, 523–535.

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., and Kenney, M. K. (**2000**). "Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex," Exp. Brain Res. **130**, 133–141.

Heinks-Maldonado, T. H., and Houde, J. F. (**2005**). "Compensatory responses to brief perturbations of speech amplitude," ARLO **6**, 131–137.

Hirano, M., and Ohala, J. (**1969**). "Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles," J. Speech Hear. Res. **12**, 362–373.

Houde, J. F., and Jordan, M. I. (**1998**). "Sensorimotor adaptation in speech production," Science **279**, 1213–1216.

Isshiki, N. (**1964**). "Regulatory mechanism of voice intensity variation," J. Speech Hear. Res. **7**, 17–29.

Jones, J. A., and Munhall, K. G. (**2002**). "The role of auditory feedback during phonation: Studies of Mandarin tone production," J. Phonetics **30**, 303–320.

Kochanski, G., Grabe, E., Coleman, J., and Rosner, B. (**2005**). "Loudness predicts prominence: Fundamental frequency lends little," J. Acoust. Soc. Am. **118**, 1038–1054.

Koyama, T., Kawasaki, M., and Ogura, J. H. (**1969**). "Mechanics of voice production. I. Regulation of vocal intensity," Laryngoscope **LXXIX**, 337–354.

Ladefoged, P., and Loeb, G. (**2002**). "Preliminary studies on respiratory activity in speech," UCLA Working Papers in Phonetics **101**, 50–60.

Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.

Lane, H., Wozniak, J., Matthies, M., Svirsky, M., and Perkell, J. (**1995**). "Phonemic resetting versus postural adjustments in the speech of cochlear implant users: An exploration of voice-onset time," J. Acoust. Soc. Am. **98**, 3096–3106.

Lane, H., Wozniak, J., Matthies, M., Svirsky, M., Perkell, J., O'Connell, M., and Manzella, J. (**1997**). "Changes in sound pressure and fundamental frequency contours following changes in hearing status," J. Acoust. Soc. Am. **101**, 2244–2252.

Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., and Hain, T. C. (**2001**). "Comparisons of voice F0 responses to pitch-shift onset and offset conditions," J. Acoust. Soc. Am. **110**, 2845–2848.

Larson, C. R., Sun, J., and Hain, T. C. (**2007**). "Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control," J. Acoust. Soc. Am. **121**, 2862–2872.

Leder, S. B., Spitzer, J. B., Milner, P., Flevaris-Phillips, C., Kirchner, J. C.,

and Richardson, F. (**1987**). "Voice intensity of prospective cochlear implant candidates and normal hearing adult males," Laryngoscope **97**, 224–227.

Liu, F., and Xu, Y. (**2005**). "Parallel encoding of focus and interrogative meaning in mandarin intonation," Phonetica **62**, 70–87.

Natke, U., Donath, T. M., and Kalveram, K. T. (**2003**). "Control of voice fundamental frequency in speaking versus singing," J. Acoust. Soc. Am. **113**, 1587–1593.

Purcell, D. W., and Munhall, K. G. (**2006**). "Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation," J. Acoust. Soc. Am. **120**, 966–977.

Siegel, G., and Pick Jr., H. L. (**1974**). "Auditory feedback in the regulation of voice," J. Acoust. Soc. Am. **56**, 1618–1624.

Siegel, G. M., and Kennard, K. L. (**1984**). "Lombard and sidetone amplification effects in normal and misarticulating children," J. Speech Hear. Res. **27**, 56–62.

Sivasankar, M., Bauer, J. J., Babu, T., and Larson, C. R. (**2005**). "Voice responses to changes in pitch of voice or tone auditory feedback," J. Acoust. Soc. Am. **117**, 850–857.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).

Svirsky, M. A., Lane, H., Perkell, J. S., and Wozniak, J. (**1992**). "Effects of short-term auditory deprivation on speech production in adult cochlear implant users," J. Acoust. Soc. Am. **92**, 1284–300.

Titze, I. R., and Sundberg, J. (**1992**). "Vocal intensity in speakers and singers," J. Acoust. Soc. Am. **91**, 2936–2946.

Turk, A. E., and Sawusch, J. R. (**1996**). "The processing of duration and intensity cues to prominence," J. Acoust. Soc. Am. **99**, 3782–3790.

Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). "Effects of noise on speech production: Acoustic and perceptual analyses," J. Acoust. Soc. Am. **84**, 917–928.

Xu, Y. (**1999**). "Effects of tone and focus on the formation and alignment of F_0 contours," J. Phonetics **27**, 55–105.

Xu, Y., Larson, C., Bauer, J., and Hain, T. (**2004**). "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," J. Acoust. Soc. Am. **116**, 1168–1178.

Xu, Y., and Xu, C. X. (**2005**). "Phonetic realization of focus in English declarative intonation," J. Phonetics **33**, 159–197.

Analogical model for mechanical vibrations in flue organ pipes inferred by independent component analysis

Enza De Lauro[a)]
Physics Department, Salerno University, S. Allende Street, 84081 Baronissi, Salerno, Italy

Salvatore De Martino[b)]
*Department of Mathematics and Informatics, Salerno University, Ponte don Melillo Street,
84084 Fisciano, Salerno, Italy*

Enrico Esposito[c)]
Department of Mechanics, Polytechnic University of Marche, Brecce Bianche Street, 60131 Ancona, Italy

Mariarosaria Falanga[d)]
Physics Department, Salerno University, S. Allende Street, 84081 Baronissi, Salerno, Italy

Enrico Primo Tomasini[e)]
Department of Mechanics, Polytechnic University of Marche, Brecce Bianche Street, 60131 Ancona, Italy

Several experiments have been performed to investigate the mechanical vibrations associated with an organ pipe. The measurements have been made by using laser Doppler vibrometry, a well-known not-invasive optical measurement technique that is very widely used in structural dynamics. The recorded signals are analyzed by using a well-established decomposition method in time domain, i.e., independent component analysis. Asymptotic dynamics methods to recognize low-dimensional dynamic system associated with this wave field is then considered. The full-toned recorded signals appear decomposed into three independent components. The independent components are nonlinear due to the fractal dimension of the attractor. These results for the mechanic vibrational field are compared with those of the acoustic one. It is interesting to note that the two fields have many common characteristics. Finally, a low-dimensional dynamic system that reproduces the main characteristics of the mechanical wave field in the time and frequency domains is introduced.
© 2007 Acoustical Society of America. [DOI: 10.1121/1.2772225]

I. INTRODUCTION

Physics underlying the sound production by musical instruments has been the object of scientific interest for many years. The main task is to develop a physical model able to reproduce both some basic mechanisms and the proper specificities of the different instruments. This work is devoted to the study of flue instruments such as organ pipes, recorders, and so on. These instruments produce self-sustained tones as a result of a nontrivial coupling between a flow and a duct. For our investigations we consider the flue organ pipes, since there is not a direct interaction between the instrument and the player that can significantly affect the tone which has to be ascribed only to the source. In these instruments, the oscillations are sustained by the interactions of a hydrodynamically unstable air jet with a sharp edge (labium). The air column within the pipe and the pipe walls vibrate as one (see, e.g., Fletcher,[1] and references therein). The commonly accepted model for flue instruments of which the organ pipe

can be considered representative is the jet-drive model (see, e.g. Refs. [2] and [1]). In previous works, we have studied the acoustic field produced by flue organ pipes.[3,4] We have proved that the acoustic signal at a given point, when the pipe is full-toned, can be decomposed into three independent components. This result follows by using a well-known decomposition method in time domain, i.e., independent component analysis (ICA). This means that the acoustic oscillations can be described as the linear superimposition of three nonlinear signals. The analysis of the phase-space dimensions associated with these independent components provides the correlation dimension, which is an estimate of the attractor dimension $d=1.2$, i.e., a limit cycle. Moreover, we have introduced a simple model that reproduces one of these independent components. Such a model is based on the discussions and suggestions from many authors who have connected limit cycles to the pipe sound production.[5–8] Previous studies indicate that the vibrations associated with the wall do not affect the tone. In particular, for the organ pipes it has been stressed that any contribution by the wall is completely masked by the air column vibrations.[9–12] Nevertheless, musicians claim to be able to recognize the differences among the sounds produced by organ pipes made of different materials, thickness, and so on. Namely, organ builders think that they can vary the pipe material according to the sound

a)Electronic mail:delauro@sa.infn.it
b)Electronic mail: demartino@sa.infn.it
c)Electronic mail:e.esposito@mm.univpm.it
d)Electronic mail: rosfal@sa.infn.it
e)Electronic mail: ep.tomasini@mm.univpm.it

TABLE I. Geometrical description of our open metal line-lead alloy organ pipes.

Pitch notation	C3	C♯3	D3	D♯3	E3	F3	F♯3	G3	G♯3	A3	A♯3	B3	C4
Length (mm)	1200	1132	1089	1000	952	899	848	800	756	713	673	636	600
Diameter (mm)	84.7	81.1	77.7	74.4	71.2	68.2	65.3	62.6	59.9	57.4	55.0	52.8	50.4
Languid diameter(mm)	69.0	67.0	64.0	61.5	58.0	56.5	54.0	51.0	48.5	47.0	44.5	43.0	41.0
Mouth width (mm)	64.6	62.4	59.4	55.2	55.0	52.5	50.5	47.5	45.0	44.2	42.6	41.0	38.2
Mouth height (mm)	16.2	15.3	14.8	14.0	13.4	13.0	12.5	11.8	11.4	11.0	10.5	9.5	9.1
Wall thickness (mm)	0.95	0.95	0.95	0.90	0.90	0.90	0.85	0.85	0.85	0.80	0.80	0.75	0.75

wanted: In the case of metal flue organ pipes usually made of a tin-lead alloy, the predominance of lead or tin should darken or brighten the tone, respectively. In this work we compare the mechanical vibration field and the acoustic one by using ICA and the reconstruction of the asymptotic dynamics to single out the analogies and the differences between the two fields. The measurements of the mechanical vibrations are made by using a noninvasive method, i.e., a laser Doppler vibrometry (LDV), while we measure the acoustic field by a network of microphones. For a detailed description of the LDV to measure mechanical vibrations see Castellini et al.[13] and Esposito et al.[14,15] In particular, LDV has already been used to study eigenmodes of the flue organ pipe and the influence of the wall vibrations on transient.[11,12] The measurements are made starting from a minimum of blowing pressure to the maximum producing the full-toned note. Finally, we show a model able to reproduce the signals associated with the mechanical vibrations.

II. EXPERIMENTAL SETUP

There are many types of organ pipes that can be classified looking at the material (e.g., wood or metal), or at the asperity (edge, reed), or at the pipe shape (cylindrical, conical, or irregular). Moreover, the pipes can be open or closed or with more complex geometrical constrains. In this paper, we study open metal (tin-lead alloy) cylindrical flue organ pipes (in 8 ft rank). We study real organ pipes under realistic operating conditions. Our organ pipes and the distribution pressure system (somière) have been built by the organ builder Olivieri of Catania, Acitrezza (the characteristics of the pipes are reported in Table I). Several experiments have been performed in the laboratories of Acoustic of Salerno University in order to detect the pipe wall vibrations. The vibrational velocity measurements are carried out by using a LDV system, a noncontact optical measurement technique. The sensor is a transducer that uses optical interferometry and electronic frequency measurements to determine the frequency shift of a laser beam reflected by a moving surface. In this experiment, the setup was composed of an OMETRON laser vibrometer VS 100, which uses a helium-neon (He–Ne) laser with a power of 1 mW. This instrument is designed to detect vibrations up to a distance of 200 m with a sensitivity of 1 μm/s and a maximum frequency of 50 kHz. We perfectly match the conditions locating the laser at 1 m from the pipes.

The mechanical vibration measurements are made considering three rings surrounding the pipe at the beginning (p1), central (p2), and end part (p3). Two points are selected

for each ring: For the first, the laser is located in front of the pipe mouth and for the second the pipe is turned 90° counterclockwise. As a consequence, six time series are available for our analysis. We perform measurements while the pipe is excited with different levels of the blowing pressure. Namely, we consider three increasing levels of the blowing pressure, until the standard condition of playing, as established by the organ builder, is fulfilled. The three levels of the blowing pressure are, respectively, 2.3 mbar (low pressure), 5.5 mbar (intermediate pressure), and 7 mbar (maximum pressure). The minimum value of blowing pressure corresponds to the minimum threshold at which we get a sound by our experimental apparatus. The pipes are fixed vertically on a proper somière, which stores the air under a controlled pressure. The acoustic field is simultaneously measured with the mechanical one. It is recorded by using six microphones placed in front of the three selected points at a distance of 20 cm. The sensors (laser and microphones) are connected to a data acquisition board with a sampling frequency of 44 100 Hz. The setup scheme is reported in Fig. 1. The measurements have not been performed in an anechoic chamber, but we have evaluated the signal to noise ratio, which is equal to 20 dB (the experiments were performed during the night). However, previous investigations of the acoustic field in an anechoic chamber have shown the same features.[3,4]

We have investigated all the 8 ft rank, but we report only the results relative to the pipe generating C4 in standard

FIG. 1. A sketch of the setup indicating the position of the laser and the microphones; p1, p2, and p3 correspond to the selected measurement points.
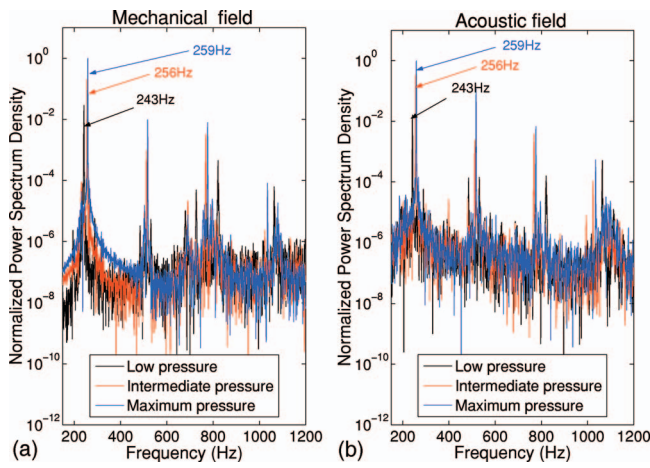
De Lauro et al.: Simple model of pipe-wall vibrations

FIG. 2. Normalized spectral density of wall (a) and acoustic (b) vibrations at different levels of blowing pressure in logarithmic scale on *y* axis. The superposition of the spectra in each case evidences the appearance of higher peaks and the trend toward the higher frequency in changing the blowing pressure from a low (black line), to intermediate (red line), and to full-toned (blue line) level.

pitch notation, looking at the stationary part of the signal. The general behavior for the other pipes is preserved. For completeness, we have made the modal analysis of the pipe. The eigenmodes of the pipe do not match exactly the acoustic field as discussed in literature (see, e.g., Kob[11,12]).

III. ANALYSIS OF THE RECORDED SIGNALS

A. Frequency domain

The general features of the mechanical wave field are evidenced looking at the signals in frequency and time domains. We start analyzing the frequency content when the pipe is excited by the different levels of the blowing pressure. We observe an enrichment in frequency content with the appearance of more peaks as the blowing pressure increases. Moreover, each frequency peak shifts, increasing its own frequency, and there is an enhancement of the relative amplitudes among the peaks: The higher peaks increase more rapidly than the fundamental one. The acoustic field displays the same behavior as shown in previous works (see, e.g., Fletcher,[1] and references therein). The power spectra of the two wave fields are sketched in Fig. 2. The superposition of the three power spectra relative to mechanical [Fig. 2(a)] and acoustic [Fig. 2(b)] fields clearly shows how the acoustic frequencies match the mechanical ones shifting in the same way while the level of the blowing pressure grows.

Namely, focusing the attention on the full-toned mechanical oscillations (Fig. 3) we observe a broadband spectrum made of many peaks. The first three correspond to those of the acoustic field. These characteristics are independent of the measure point. As in the acoustic case, we observe a broadband spectrum and an enhancement in amplitude with a shift in frequency as a function of the blowing pressure level. In particular, we note that the intermediate case is already characterized by a broadband spectrum with three predominant peaks, but we reach the right value of frequency only for blowing pressure level corresponding to the full-toned pipe.
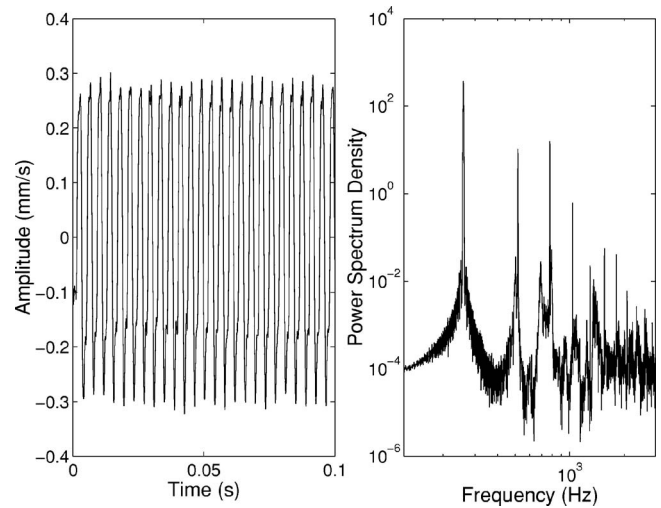
FIG. 3. Wall vibration time series and its power spectrum density (loglog scale) when the pipe is full-toned.

B. Amplitude analysis

We look at the amplitude of the mechanical vibrations along the three points of the pipe under study considering the different levels of the blowing pressure. Independent of the blowing pressure level, the velocity amplitudes are higher in the central part of the pipe. In particular, when the pipe is played with the maximum blowing pressure, the velocity amplitude is about 0.6 mm/s in the central part whereas at the beginning and at the end part is about 0.18 mm/s (see Fig. 4).

It is interesting to look at the ratio among the amplitudes of the frequency peaks at the three rings. At the beginning ring of the pipe, we observe that the amplitude of the fundamental is about 5 times the second and the third. At the center, we observe that the first peak is about 25 times the second peak and 12 times the third peak. At the end part, we observe that the fundamental peak is about 12 times the second and third one. Hence, the higher amplitude is associated with the fundamental peak at the central point of the pipe. This characteristic behavior of the mechanical vibrations
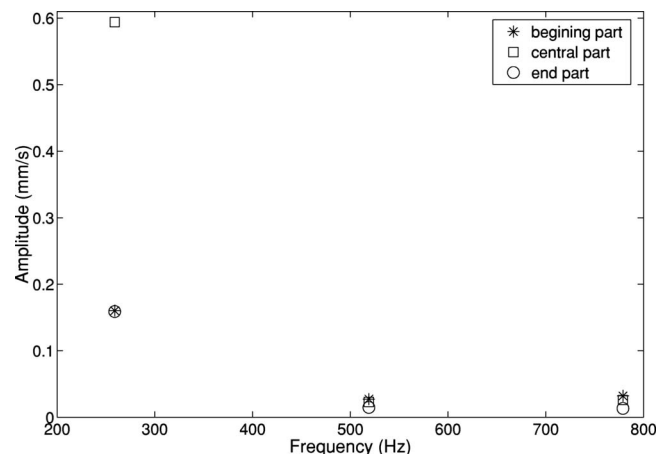

FIG. 4. Velocity amplitudes vs frequency of the mechanical vibrations of the full-toned pipe relative to the beginning (asterisk), central (square), and end point (circle) of the pipe.
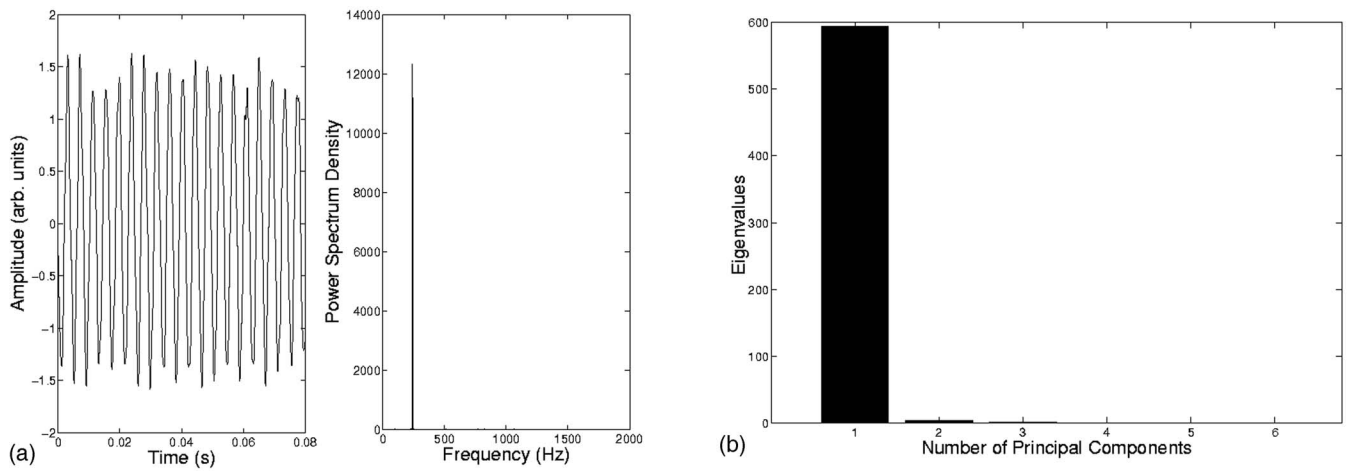
FIG. 5. (a) Extracted component by ICA and its power density spectrum obtained analyzing the mechanical vibrations when the pipe is excited with the minimum level of the blowing pressure. The component corresponds to the frequency peak at 243 Hz. (b) PCA evidences one principal eigenvalue.

does not correspond to the recorded acoustic field where the ratio among the amplitudes of the peaks does not depend on the point considered.

C. Time-domain analysis

1. Method: Independent component analysis

In Secs. II A and II B, looking at the spectral properties, we have shown the rising and the shift of the frequencies at increasing blowing pressure for the mechanical field as well as for the acoustic field. Due to the nonlinearity of the problem we are dealing with, frequency methods have to be supported by time-domain techniques. In particular, we adopt a specific time-domain technique, i.e., ICA, which was already applied to the acoustic field by Bottiglieri et al.[3] In that case, the application of ICA showed that the acoustic field could be decomposed into nonlinear independent components.

ICA is an entropy-based technique, which can find underlying factors or components from multivariate (multidimensional) statistical data on the basis of their statistical independence evaluated using fourth-order statistical properties. Though ICA has been well known from the begining of the 1980s, it became very popular after the introduction of FastICA, for its robustness and rapid convergence. Let us explain in brief the idea on which ICA is based. We suppose to have m different recorded time series $\mathbf{x}$ that we hypothesize to be the linear superposition of n mutually independent unknown sources $\mathbf{s}$, due to different mixing, represented by a constant unknown matrix $\mathbf{A}$. This mixing is essentially due to path, noise, instrumental transfer-functions, etc. The recorded series are the instantaneous linear mixtures of some independent dynamical systems. If the mixing has to be linear, nothing is assumed with respect to the sources, which can be linear or nonlinear. Under these hypotheses, the method extracts the independent components in the time domain (for more details see Hyvarinen et al.,[16] and many papers cited therein; the application to dynamic systems is considered in particular in De Lauro et al.[17]).

This method cannot determine the variances (proportional to energies) of the extracted independent components. Therefore, the most natural way to obtain a unique expansion is to assume that each source has a unit variance. This is accomplished by principal component analysis (PCA). This method involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for the remaining variability. Traditionally, PCA is accomplished by diagonalization of covariance matrix.[18]

2. Results

We perform ICA considering the recorded signals at the three increasing levels of the blowing pressure. For each of the three selected levels of the blowing pressure, we consider as different linear mixtures the registrations of the mechanical vibrations relative to the points described in Sec. II. Hence, for each level of the blowing pressure we have six mixtures as input of ICA.

We obtain the following results:

(1) For the lowest level of the blowing pressure, we have just one extracted component in the time domain corresponding to a frequency of 243 Hz [Fig. 5(a)]. The diagonalization of the covariance matrix (PCA) of the recorded signals (which is a linear technique) evidences two principal eigenvalues as indicated in Fig. 5(b), nevertheless only one is recognizable as an independent component.

(2) For the intermediate and the highest levels of the blowing pressure, ICA extracts three signals in the time domain as independent components (Figs. 6 and 7). For the intermediate level of pressure, the extracted component are peaked, respectively, at 256, 512, and 768 Hz [Fig. 6(a)]. PCA clearly evidences three eigenvalues as shown in Fig. 6(b). Finally, when the organ pipe is full-toned, ICA recognizes three independent components at the frequencies equal to 259, 518, and 777 Hz, respectively [Fig. 7(a)]. These frequencies correspond to the main peaks in the origi-
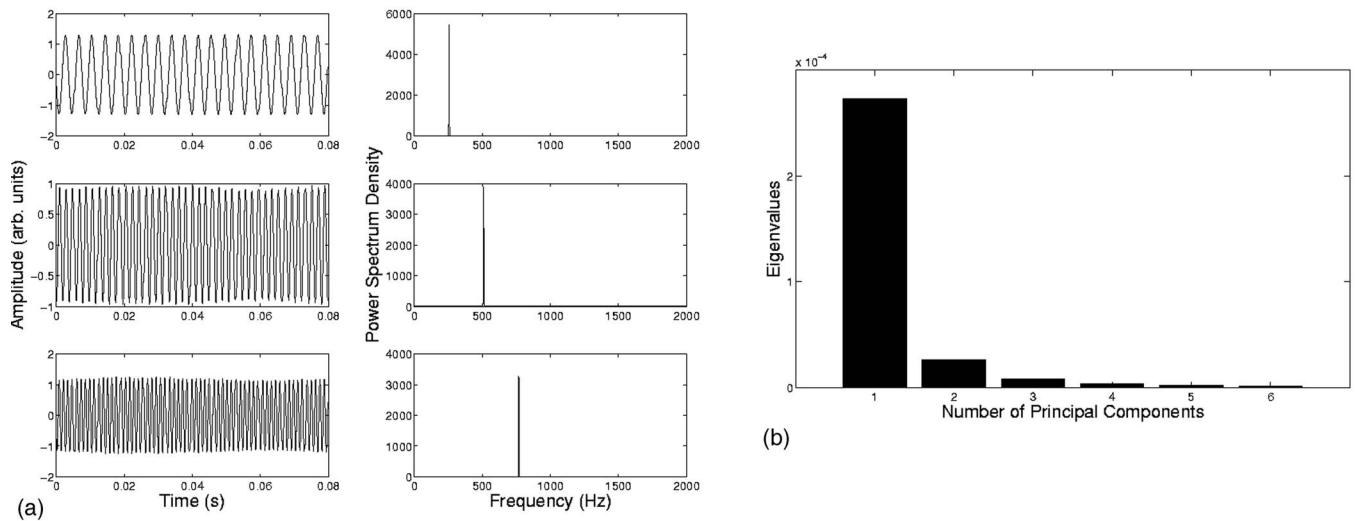
FIG. 6. (a) Extracted components by ICA and their spectra obtained analyzing the mechanical vibrations when the pipe is excited with intermediate level of the blowing pressure. The components correspond, respectively, to the frequency peaks at 256 and 512 Hz; (b) three principal eigenvalues are significant.

nal broadband spectrum of Fig. 2. Once again, three principal components are evidenced [Fig. 7(b)].

(3) The acoustic field simultaneously measured displays the same behavior for each level of the blowing pressure. The extracted wave forms for the two fields are similar, as confirmed by the very high correlation coefficients (greater than 0.95).

Hence, the mechanical vibration wave field, as in the acoustic case, can be decomposed into independent signals in time. Before introducing a phenomenological model we consider another step to be necessary, i.e., to get some information about the dynamic system underlying the extracted components.

IV. DETERMINATION OF THE LOW-DIMENSIONAL DYNAMIC SYSTEM

A. The method

A standard way to proceed when dealing with time series is to reconstruct by the time delay method the phase space where the asymptotic dynamics evolves. This leads to establishing the degrees of freedom necessary to fully get all the main information about the dynamics. This method relies on the mathematical formulation due to Takens.[19] If s is a time series of n scalar observations sampled at equal intervals (pipe-wall vibrations), the reconstructed attractor consists of points of the form $x_i = (s_i, s_{i+\tau}, \ldots, s_{i+(m-1)\tau})$, where m is the embedding dimension and τ is the time delay. Takens shows, under suitable hypotheses, that this reconstruction is equivalent to the original attractor if $m \geq 2d_a + 1$, where d_a is the attractor dimension. Thus the numerical problem is to determine m and τ. There is a lot of literature on this argument (see, for example, Abarbanel[20]). Among several but essentially equivalent methods, we select average mutual information (AMI)[21] and false nearest neighbors (FNN)[22] to calculate, respectively, time delay and embedding dimension. AMI is the nonlinear extension of the autocorrelation function to the nonlinear domain. FNN is based on the projection of the points of the dynamics onto spaces of increasing embedding dimensions. The points appear nearest neighbors of
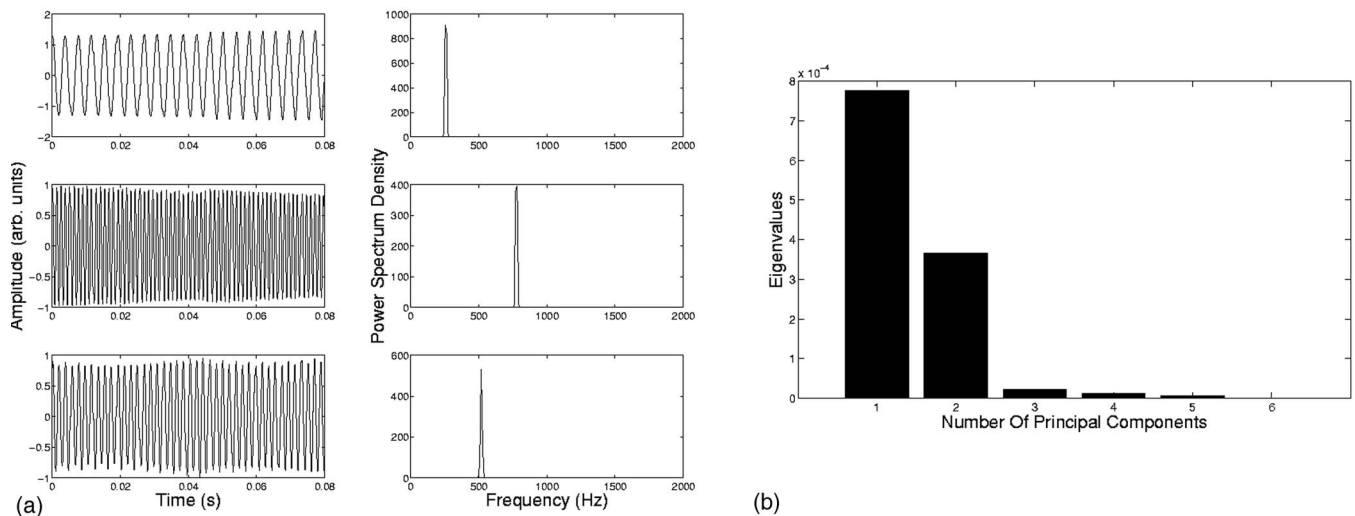

FIG. 7. (a) Extracted components by ICA and their spectra obtained analyzing the mechanical vibrations in the full-toned pipe. The components correspond, respectively, to the frequency peaks at 259, 518, and 777 Hz; (b) three principal eigenvalues are relevant.
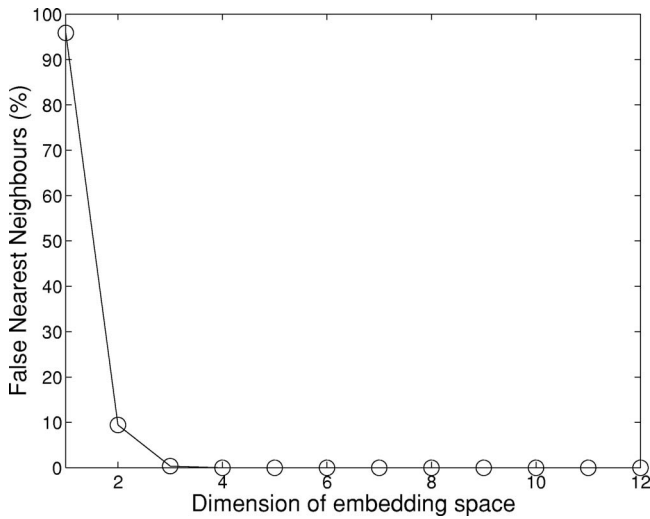
FIG. 8. Application of FNN to the mechanical vibration when the pipe is full-toned. The fraction of false nearest neighbors as a function of the embedding dimension for a time window of 1 s. We obtain as upper bound $m=4$.
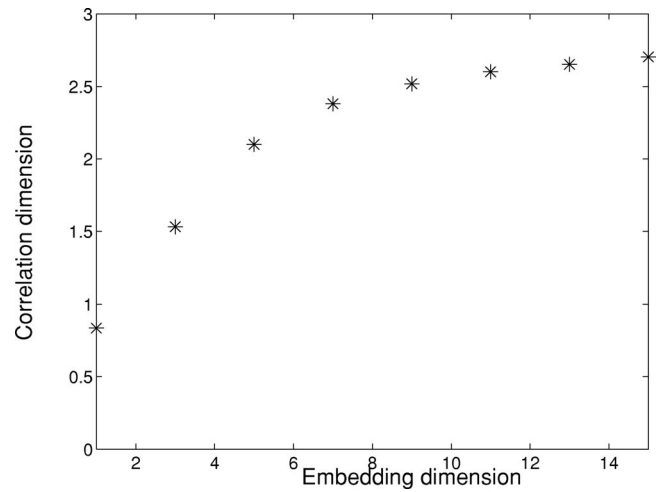

FIG. 9. Application of Grassberger and Procaccia to the mechanical vibration when the pipe is full-toned: correlation dimension vs the embedding dimension for a time window 1 s long. The plateau is stable around 2.8, suggesting that the entire signal is low dimensional.

some others until the dimension becomes the proper one of our embedding space. When the fraction of these false nearest neighbors with respect to the total number of points is zero then we can stop our process. Note that the value of m, so obtained, represents the lower limit for the embedding dimension necessary to evaluate attractor dimension d_a on which the asymptotic dynamics evolves. Moreover, it is an upper bound for d_a. In order to estimate d_a, we use the standard technique of Grassberger and Procaccia[23] based on the calculation of the correlation integral:

$$C(l) = \frac{1}{N} \sum_i \frac{1}{N-1} \sum_{j \neq i} \vartheta(l - |x_i - x_j|), \qquad (1)$$

where x_i are the vectors previously defined, and ϑ is the Heaviside function. The slope of $C(l)$ in the scaling region (where it has a power law behavior) gives the correlation dimension, which is an estimate of the dimension of the attractor.

B. The results

We consider the pipe-wall vibrations relative to the three different points at the highest level of pressure. First, we estimate the embedding dimension (m) via FNN. The percentage of FNN becomes zero for the embedding dimension equal to four. This suggests that the dynamic can be embedded in a four-dimensional space. In Fig. 8, we report, the results of a time window of recording 1 s long relative to the point in the central part of the pipe.

The technique for the other points and for different length series shows the same result. The result is refined applying the algorithm of Grassberger and Procaccia to evaluate the attractor dimension. In Fig. 9 we show the correlation dimension of the mechanical vibration series versus the embedding dimension. The existence of the plateau suggests that the system is deterministic, nonlinear, and characterized by a correlation dimension equal to 2.8. This states that the nonlinear dynamic system associated with the

mechanism that induces the mechanical vibrations is low dimensional. Analyzing the acoustic field, Bottiglieri et $al.$[3] have individuated the same value for the phase space dimension.

If we apply this technique to the signals, filtered in the frequency band identified by ICA, the embedding dimension and then the attractor dimension decrease. As before, a first estimate of the embedding dimension is furnished by FNN. For each frequency band, the percentage of FNN is zero for an embedding dimension equal to 2, as one can see in Fig. 10.

Finally, the estimates of the correlation dimensions for each band are contained in Fig. 11, where the existence of a plateau characterized by a dimension less than 1.5 for all the signals indicates a very low-dimension dynamical system.
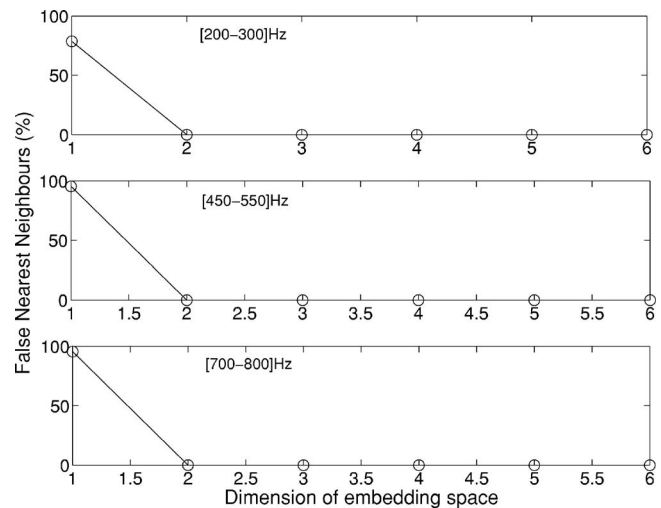

FIG. 10. Application of FNN to the mechanical vibrations, when the pipe is full-toned, filtered in the frequency bands individuated by ICA: [200–300], [450–550], and [700–800] Hz. The percentage of false nearest neighbors as a function of the embedding dimension goes to zero for embedding dimension equal to 2.

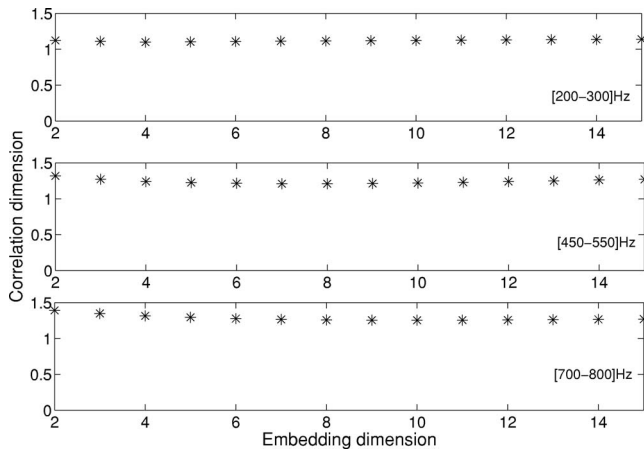De Lauro et $al.$: Simple model of pipe-wall vibrations

FIG. 11. Correlation dimension vs the embedding dimension for the mechanical vibrations filtered in the frequency bands extracted by ICA: [200–300], [450–550], and [700–800] Hz (full-toned pipe).

V. SUMMARY OF EXPERIMENTAL RESULTS

We can summarize our investigations as follows:

(1) Looking at the behavior of the amplitudes along the pipe in the considered rings (p1,p2,p3), we observe highest amplitude oscillations along the ring p2 and lowest amplitude oscillations along the rings p1 and p3 independent of the frequency.

(2) Taking a frequency content at growing the blowing pressures, we observe, as for the acoustic field, the appearance of several peaks and the shift in frequency at higher pressure.

(3) The application of ICA gives as a result, as pressure increases, one component and three components: one component at a certain frequency for the lowest level of the blowing pressure; three components with certain frequencies and suitable amplitude ratios for the intermediate level of the blowing pressure; at the highest level of pressure three components with frequencies corresponding to those of the full-toned note. This behavior suggests that we have linearly coupled components nonlinearly produced.

(4) Our system is characterized by a low-dimensional dynamic system. If it is linear, i.e., composed of normal modes, we should observe as phase-space dimension $d=3$. If we estimate the phase space dimension by using standard methods of analysis such as mutual information, FNN, and the Grassberger and Procaccia method, we obtain $d=2.8$, for the full toned recorded signal and $d=1.2$ and for the single component. This is a clear indication of nonlinearity. Notice that these dimensions coincide with those obtained for the acoustic field.[3]

These four points still hold for the acoustic field, except for point 1. Namely, the amplitudes of the acoustic components are independent of the points along the pipe. We would like to remark that if we look at the spectrum we can observe a continuous development as the blowing pressure increases, but ICA gives only one component at the lowest level of blowing pressure and three components at the intermediate

and highest levels. It would be misleading to give the impression that a frequency domain view is never useful: Self-sustained musical instruments exhibit behaviors that can be illuminated from this point of view, but it is misleading to sustain, on the contrary, that for nonlinear signals as self-oscillations is not important to adopt, at the same time, approaches based on time domain. ICA looks at signals in time domain and establishes that there are many independent components. If we look at one of these independent components in Fourier transform we observe several peaks, which are connected with the first as higher harmonics. It is well known from the theory of signals that limit cycles require infinite Fourier components to be reproduced, though the time signals have a characteristic wave form in which only one well-defined period is recognizable. Our results suggest that three suitable coupled nonlinear dynamical systems can reproduce the main features of the mechanical vibrations.

VI. MODEL

A. Analogical model

The goal of the physics of the organ pipe is to individuate a mechanism essentially hydrodynamic, which is able to produce "oscillating" signals (i.e., acoustic or vibrational fields) by a constant source (pressure reservoir) without implying any periodic forcing. The general theory must provide the hydrodynamic equation coupled with an equation of the elastic system (body pipe). In literature there are many proposed models that look at a particular aspect of the organ pipe system or at the system as a whole. These models, by using suitable approximations, reproduce the characteristic behavior of full-toned organ pipe looking essentially at the frequency domain. We chose a different approach, namely our main aim is to obtain a macroscopical model, i.e., a set of ordinary differential equations able to reproduce the characteristic behavior in the frequency and time domains, i.e., the enhancement in the observed spectra at the increasing blowing pressure and the independent components in time. Our starting point is the same as widely accepted in literature—that the appearance of a globally organized phenomenon, due to a confined hydrodynamical jet, is produced by a feedback effect. The set of equations of the triode is a well-known example of a system that produces globally organized oscillations.[5] The electric circuit consists of a triode, which plays the role of a constant source, and a RLC circuit coupled with the triode, which resonates. The feedback mechanism consists in changing the grid potential through a coupling element with the anodic current.

This physical system generates self-oscillating signals and the related macroscopic equations well reproduce the experimental signals. We want to write a similar set of equations for the organ pipe body searching for the most simple system able to reproduce our observations. In this analogy, we must identify, even if in an heuristic way, structures that have a similar role in the vacuum tube and in the organ pipe. The feedback mechanism will be selected on its capacity to reproduce our results.

B. The Andronov oscillator

We briefly describe the triode circuit. We are considering the basic electric circuit, i.e., a triode self-coupled through a RLC circuit[24] where the coupling is made through an inductance. Obviously, the solutions of the equations are very difficult to obtain analytically. In fact, we have three coupled equations, one of which is the characteristic equation, which links the differential anodic current to the differential grid and cathode electric potentials. Typically, people approximate this characteristic equation by its solution obtained phenomenologically, i.e., the characteristic function. Among the possible phenomenological characteristic functions (sigmoid, hyperbolic function, etc.), the most used is the piecewise linear approximation. The general equations for this system called Andronov oscillator are

$$L\ddot{I} + R\dot{I} + \frac{1}{C}I = \frac{I_a}{C},$$

$$I_a = f(u_g),$$

$$u_g = -M\dot{I}, \tag{2}$$

where I represents the current in the RLC circuit, I_a is the anodic current, u_g is the grid potential, M is the mutual inductance (which has to be negative to install self-coupling), and $I_a = f(u_g)$ is the characteristic function. Choosing the piecewise linear approximation as a characteristic function, the previous equations become in a-dimensional variables:

$$\ddot{x}_1 + h_1\dot{x}_1 + \omega_1^2 x_1 = 0 \quad x < b,$$

$$\ddot{x}_1 - h_2\dot{x}_1 + \omega_1^2 x_1 = 0 \quad x > b. \tag{3}$$

In this way, the nonlinearity is concentrated in the threshold b, which separates the phase space into two regions; h_1 takes into account the dissipation; h_2 and b contain the coupling; ω_1 is the natural angular frequency. Numerically integrated, this equation generates a self-sustained current. Though the piecewise linear approximation is very crude, the generated signal retains all the characteristics of the experimental current. This signal shows a Fourier transform with infinite components as must be for nonlinear signals. We note that limit cycle is the simplest dynamical attractor, obviously not chaotic.

C. Adaptation of the Andronov model to organ pipe

Having shown that when the pipe is full-toned only three independent components are excited and they are in linear combination, we assume that the pipe can be represented by three linearly coupled RLC circuits (i.e., capacitive coupling). One of these circuits is coupled with the triode grid by the mutual inductance. Our equations are

$$L_1\ddot{I}_1 + R_1\dot{I}_1 + \frac{1}{C_1}I_1 + \frac{1}{C_{12}}(I_1 - I_2) = \frac{I_a}{C_1},$$

$$L_2\ddot{I}_2 + R_2\dot{I}_2 + \frac{1}{C_2}I_2 + \frac{1}{C_{12}}(I_2 - I_1) + \frac{1}{C_{23}}(I_2 - I_3) = 0,$$

$$L_3\ddot{I}_3 + R_3\dot{I}_3 + \frac{1}{C_3}I_3 + \frac{1}{C_{23}}(I_3 - I_2) = 0,$$

$$I_a = f(\dot{I}_1), \tag{4}$$

where I_i is the current in each circuit, R_i is one of the electric resistances, C_i is the capacity of each circuits, C_{ij} are the coupling capacities between the circuits, and $I_a = f(\dot{I}_1)$ is the piecewise linear characteristic function. Explicating the characteristic function, Eq. (4) becomes

$$L_1\ddot{I}_1 + R_1\dot{I}_1 + \frac{1}{C_1}I_1 + \frac{1}{C_{12}}(I_1 - I_2) = 0, \quad I_1 < I_b,$$

$$L_1\ddot{I}_1 - \gamma_1\dot{I}_1 + \frac{1}{C_1}I_1 + \frac{1}{C_{12}}(I_1 - I_2) = 0, \quad I_1 > I_b,$$

$$L_2\ddot{I}_2 + R_2\dot{I}_2 + \frac{1}{C_2}I_2 + \frac{1}{C_{12}}(I_2 - I_1) + \frac{1}{C_{23}}(I_2 - I_3) = 0,$$

$$L_3\ddot{I}_3 + R_3\dot{I}_3 + \frac{1}{C_3}I_3 + \frac{1}{C_{23}}(I_3 - I_2) = 0. \tag{5}$$

In this system the pumping is concentrated in the parameter γ_1, which regards only the first circuit, i.e., what happens at the edge. The first couple of equations give the nonlinearity of the system, generating a limit cycle under certain conditions, but the propagation of these self-oscillations is linearly transmitted to the other circuits. To be clear, for us the triode is the constant source, the RLCs circuits describe in a suitable way the organ pipe body; the characteristic function represents the relation between flow and pressure at the edge. In fact, in this way we are approximating the feedback mechanism. The equivalence between the elements of electric circuits and corresponding mechanical parameters are well known.

We have to fix the parameters L_i, C_i, and C_{ij} in order to get the first component (ω_0), the second ($2\omega_0$), and the third ($3\omega_0$). Different possible choices can be taken for the parameters. We assume that $L_1 = L$, $L_2 = L/2$, $L_3 = L/3$, this corresponds to discretizing the pipe body looking at oscillating masses associated with the three observed independent components. Moreover, we assume that $C_1 = C_2 = C_3 = \gamma C$, which correspond to assuming that the elasticity parameters are the same. Finally, the coupling capacities $C_{12} = \alpha C$ and $C_{23} = \beta C$ are fixed taking into account the frictionless linearized counterpart of the system, i.e., with $I_a = 0$ everywhere. This linearized system can be diagonalized and this brings a particular choice for γ, α, and β, i.e., $\gamma = 2/3$, $\alpha = 5/3$, and $\beta = 1$. Going back to the nonlinear system, it can be written in dimensionless variables as follows:

$$\ddot{x}_1 + h_1\dot{x}_1 + \frac{2}{3}\omega_0^2 x_1 + \frac{5}{3}\omega_0^2(x_1 - x_2) = 0, \quad x_1 < b,$$

$$\ddot{x}_1 - \gamma_1\dot{x}_1 + \frac{2}{3}\omega_0^2 x_1 + \frac{5}{3}\omega_0^2(x_1 - x_2) = 0, \quad x_1 > b$$

$$\ddot{x}_2 + h_2\dot{x}_2 + \frac{4}{3}\omega_0^2 x_2 + \frac{10}{3}\omega_0^2(x_2 - x_1) + 2\omega_0^2(x_2 - x_3) = 0,$$
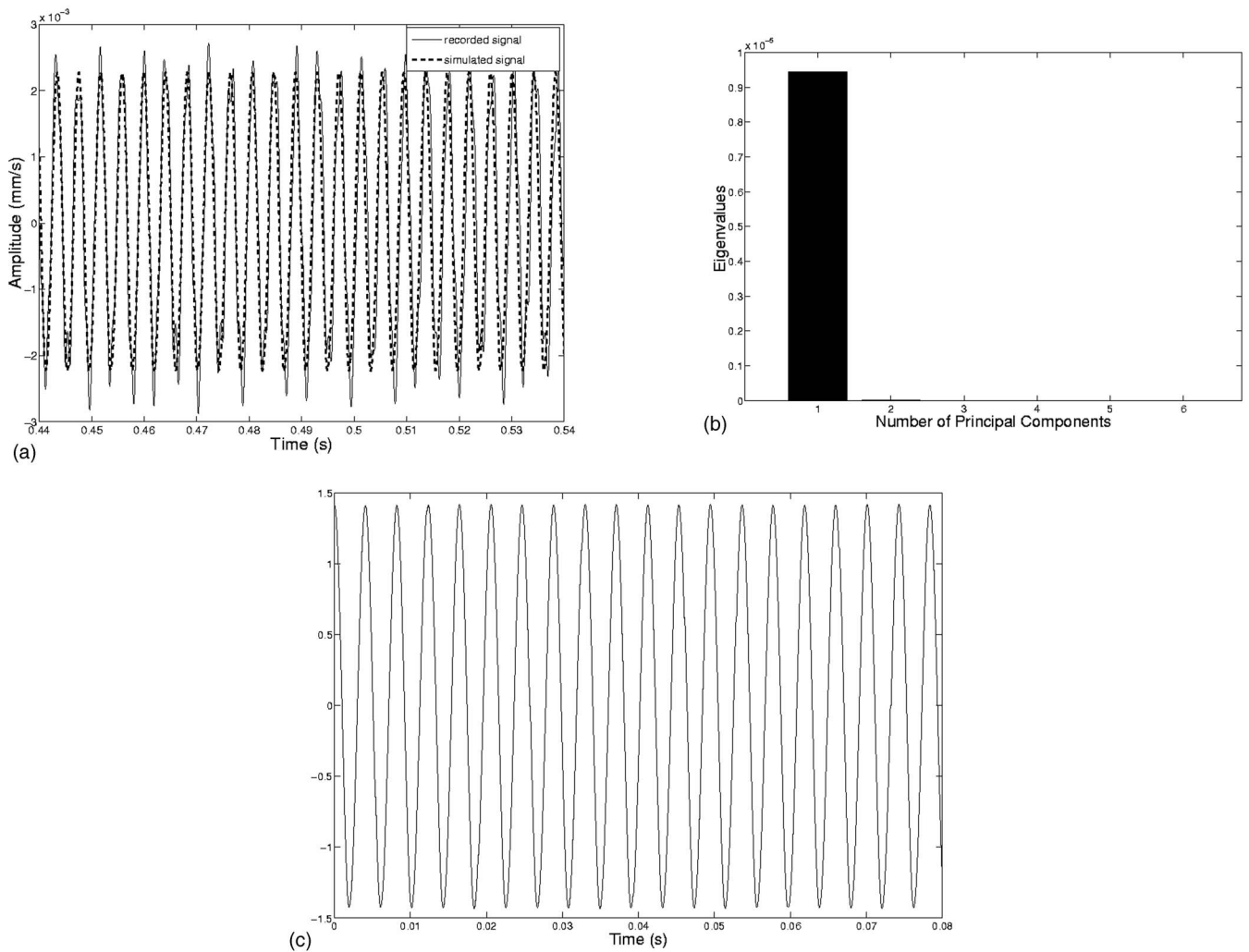
FIG. 12. (a) Comparison between recorded and simulated signals with a minimum value of pumping: $h_1 = 1200$, $\gamma_1 = h_2 = h_3 = 280$, $b = -0.002$, $x_1 = 0.01$, $x_2 = x_3 = \dot{x}_1 = \dot{x}_2 = \dot{x}_3 = 0$; (b) PCA gives just one eigenvalue, whose wave form is extracted by ICA and reported in (c). The frequency content of the simulated signal is plotted in Fig. 15 with the black line.

$$\ddot{x}_3 + h_3\dot{x}_3 + 2\omega_0^2 x_3 + 3\omega_0^2(x_3 - x_2) = 0. \qquad (6)$$

In this system; ω_0 is the angular frequency related to the fundamental eigenmode; the free parameters are: b, which represents the minimum flux which gives the onset of the oscillations; γ_1, which is the pumping parameter; and h_i's, which are linked to the dissipations in the pipe which may change according to γ_1.

In Sec. VI D we numerically integrate this equation with different and increasing values of pumping and compare the results of this simulation with the experimental recorded signals of the mechanical vibrations.

D. Numerical integration of the analogical model

Now we integrate Eq. (6) to match the numerical solutions with the recorded note. As a first step we want to obtain a simulated signal that reproduces the observed vibrations when the pipe is blown with the minimum value of pressure. The best fit with the experimental data brings to fix $b = -0.002$, $h_1 = 1200$, and $\gamma_1 = h_2 = h_3 = 280$. The comparison between the simulated and the recorded signal is reported in Fig. 12(a). The application of ICA to the simulated note pro-

vides one extracted component with the same frequency of the recorded signal (243 Hz) [Fig. 12(c)]. PCA analysis performed on the simulated signals indicates just one principal eigenvalue [Fig. 12(b)].

To get the recorded note at the intermediate level of the blowing pressure, we do not change h_1, but it is sufficient to increase γ_1 (and accordingly h_2 and h_3) to the value 480. This provides a simulated signal with the same amplitude and the same shift in frequency observed in the recorded signals [Fig. 13(a)]. Moreover, the application of ICA of the simulated signal extracts three components related to the three principal components shown in Fig. 13(b) as in the case of the recorded signals [Fig. 13(c)]. The value of b is fixed to -0.011.

Finally, without varying h_1 and b, we grow the value of γ_1 until 620 obtaining a simulated signal with high correlation with the recorded signal when the pipe is blown with maximum level of pressure. Applying ICA we obtain once again three components but with the characteristic frequencies of full-toned note (Fig. 14).

Summarizing, our system is able to reproduce the enhancement in amplitude and the shift in frequency increasing
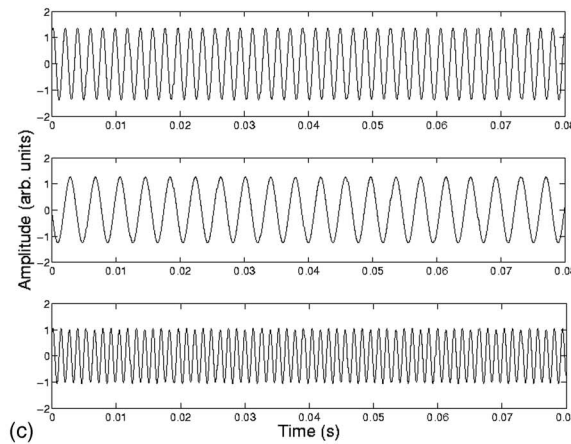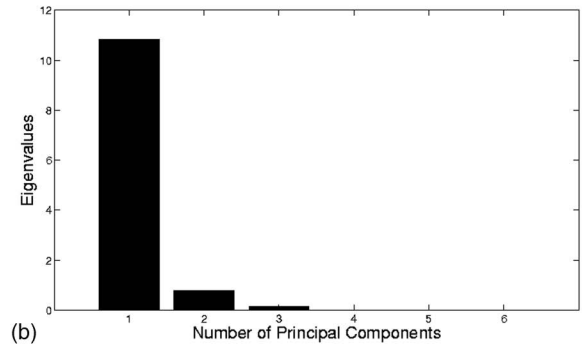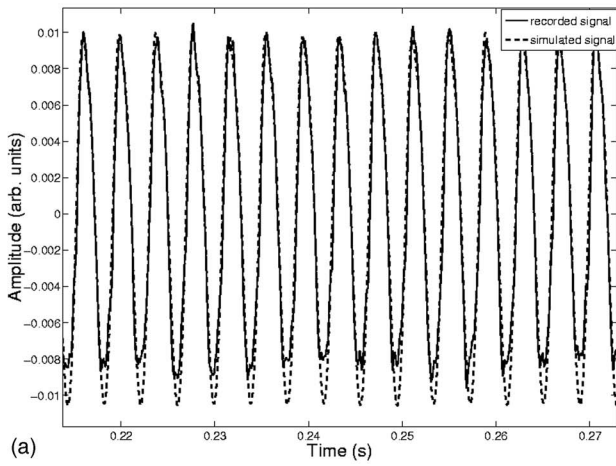
FIG. 13. (a) Comparison between recorded and simulated signals with the intermediate value of pumping: $h_1 = 1200$, $\gamma_1 = h_2 = h_3 = 480$, $b = -0.011$ with the same initial conditions as in Fig. 12; (b) PCA gives three eigenvalues, whose wave forms are extracted by ICA and reported in (c). The frequency content is plotted in Fig. 15 with the red line.

the pumping parameter as in the real case. This is clearly shown in Fig. 15 where the frequency content of the simulated signals at the different levels of pumping is reported. The application of ICA to the simulated note provides the same results as for the recorded ones, i.e., one extracted component for the lowest value of pumping and three components for the other two. The system of Eq. (6) provides another state: Changing the pumping we can induce the excitation of only two components associated with the first and the second extracted components. In the experimental case, we could not vary the pressure level continuously and this prevented us from checking this situation (we passed from one to three). Applying the analysis of Grassberger and Procaccia we obtain for the simulated system the same fractal dimensions of the recorded signals.

VII. CONCLUSION

We have analyzed the mechanical vibrations of the body of an organ pipe (C4) in stationary regime by using LVD. We have shown that they are decomposed into independent components until three when the pipe is fully played. This means that the signals recorded can be written as a linear combination of independent signals with a suitable amplitude. The

analysis of signals with standard methodology to infer phase-space dimensions shows fractal dimension for the whole signal and the independent components. In other words we have a linear combination of nonlinear signals.

These characteristics have already been observed for the acoustic field.[3] The amplitudes of the recorded signals for the mechanical vibration and the acoustic fields show different behaviors.

It is well known that the mechanical and the acoustic fields appear to be globally organized oscillations induced by a confined jet interacting with the sharp edge. This interaction affects the flow. As a consequence, a feedback mechanism produces a limit cycle.[25] In other words, both fields are an effect of a globally organized phenomenon that appears in confined geometry due to a feedback mechanism.

Models relative to self-oscillations produced by organ pipe that consider feedback mechanisms looking at pressure-flow characteristics are already discussed in literature (see, e.g., Hirschberg et al.[26]). In our paper we have introduced an analogical model inserting all the experimental results we have obtained, looking at the mechanical vibrations. With respect to those models we consider a more complex structure describing the pipe body, namely more elements linearly coupled. We are considering a macroscopic model to repro-
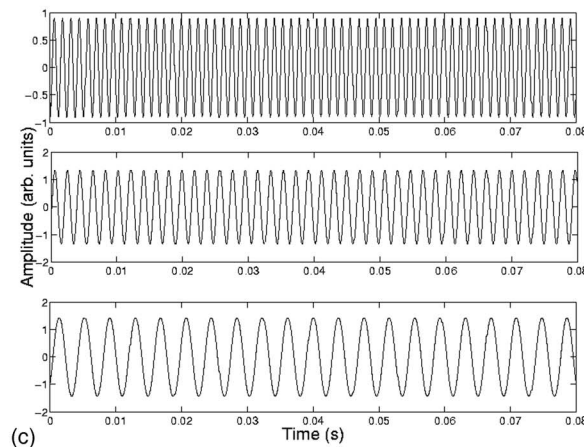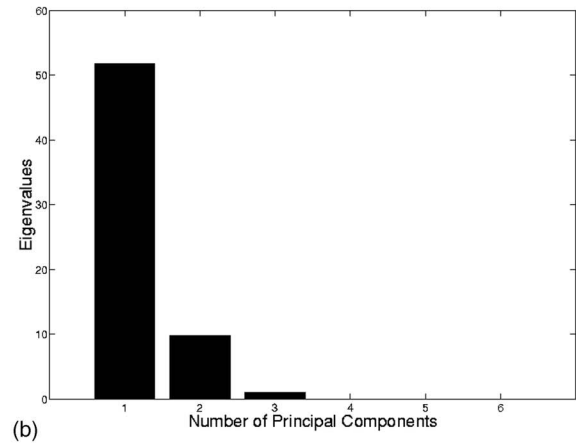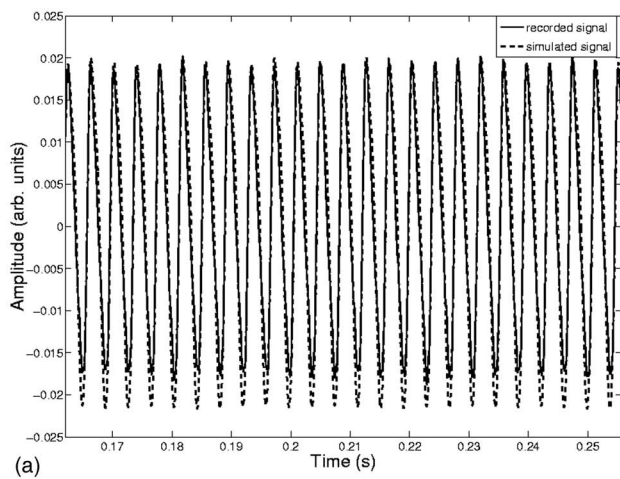
FIG. 14. (a) Comparison between recorded and simulated signals with the maximum value of pumping: $h_1=1200$, $\gamma_1=h_2=h_3=620$, $b=-0.011$ with the same initial conditions as in Fig. 12; (b) PCA gives three eigenvalues, whose wave forms are extracted by ICA and reported in (c). The frequency content is plotted in Fig. 15 with the black line.

duce the behavior of the body pipe at a point. Moreover, with the experience in vacuum tube systems rather than considering a very complex structure for the pressure-flow characteristics, we extract a linear one. As a consequence, our system reduces the nonlinearity linked to the feedback mechanism to

FIG. 15. Normalized power spectrum density of the simulated signals [Eq. (6)] in logarithmic scale on y axis. As one can see, the shift in frequency and the enhancement in amplitudes are well reproduced.

consider equation piecewise linear. As in the triode case for the self-current, it is interesting to note that a so simplified model is able to reproduce not only the experimental full toned signals but also, varying a pumping parameter, the observed phenomenon at increasing blowing pressure. Numerical simulation allows one to change continuously the pumping parameter. Consequently the model shows many states with one independent component, many with two components, and finally many with three independent components which become, at a suitable value of the pumping, exactly those of the full-tone tone. It is intriguing that if we adjust parameters looking at acoustical signals, this analogical model can also give the acoustic signals. This does not mean that the acoustical field is generated by pipe body vibrations, since it is well known that the energy associated with the mechanical vibrations is not sufficient for that aim as is shown also by our analysis on the amplitudes.

As one can see, in a recent movie obtained looking inside the section of pipe where the edge is located,[27] the oscillating confined jet appears as the source of globally organized oscillations of the pipe body and of the fluid contained in the pipe. The acoustical field should be a suitable superposition of these two effects.
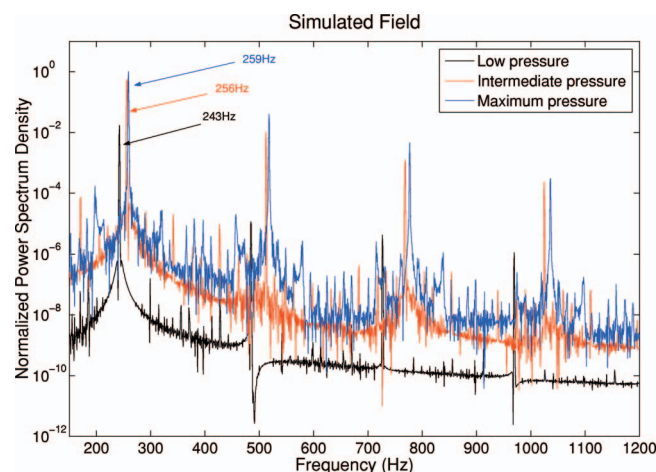
[1]N. H. Fletcher, "The nonlinear physics of musical instruments," Rep. Prog. Phys. **62**, 723–764 (1999).

[2]J. W. Coltman, "Jet drive mechanism in edge tones and organ pipes," J. Acoust. Soc. Am. **60**, 725–733 (1976).

[3]M. Bottiglieri, E. De Lauro, S. De Martino, and M. Falanga, "Analogical model inferred from time domain analysis and generating organ pipes self sustained-tone," Int. J. Mod. Phys. B **18**, 509–518 (2004).

[4]E. De Lauro, S. De Martino, M. Falanga, and G. Sarno, "Constraints on the model for self-sustained sounds generated by organ pipe inferred by independent components analysis," *Proceedings of Sound and Music Computing '05 XV CIM*, Salerno, Italy, 24–26 November 2005.

[5]A. A. Andronov, "Poincaré limit cycles and the theory of self-sustaining oscillations," Comptes Rendus de l'Académie des Sciences de Paris **189**, 559-561 (1929).

[6]H. Hirschberg, "Aero-acoustics of wind instruments," in *Mechanics of Musical Instruments*, CISM Lecture Series (Springer, Berlin, 1996).

[7]J. Woodhouse, "Self-sustained musical oscillators," in *Mechanics of Musical Instruments*, CISM Lecture Series (Springer, Berlin, 1996).

[8]S. Yoshikawa and J. Saneyoshi, "Feedback excitation mechanism in organ pipes," J. Acoust. Soc. Jpn. (E) **1** 3, 175–191 (1979).

[9]J. Backus and T. C. Hundley, "Wall vibrations in flue organ pipes," J. Acoust. Soc. Am. **39**, 936–945 (1966).

[10]C. P. Boner and R. B. Newman, "The effect of wall materials on the steady-state acoustic spectrum of flue pipes," J. Acoust. Soc. Am. **12**, 83–89 (1940).

[11]M. Kob, "Influence of wall vibrations on the transient sound of a flue organ pipe," Acust. Acta Acust. **86**, 642–648 (2000).

[12]M. Kob, "Eigenmodes of a flue organ pipe," Acust. Acta Acust. **86**, 755–757 (2000).

[13]P. Castellini, G. M. Revel, and E. P. Tomasini, "Laser Doppler Vibrometry: A review of advances and applications," *The Shock and Vibration Digest* (Sage, 1998), Vol. 30, pp. 443–456.

[14]E. Esposito, M. Navarri, S. Papalini, M. Pontillo, E. P. Tomasini, and R. Toppi, "Experimental determination of air influence on loudspeaker cone vibrations by Scanning Laser Doppler Vibrometry," *Proceedings of the fifth International Conference on Vibration Measurements by Laser Techniques*, 18–21 June 2002, SPIE, Bellingham, WA, Vol. **4827**, pp. 447–457.

[15]E. Esposito, "A comparative study of the vibro-acoustical behaviour of electric guitars produced in different decades," *Proceedings of the Stockholm Music Acoustics Conference 2003, SMAC 2003*, Stockholm, 6–9 August 2003, pp. 125–128.

[16]A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis* (Wiley & Son, Inc., New York, 2001).

[17]E. De Lauro, S. De Martino, M. Falanga, A. Ciaramella, and R. Tagliaferri, "Complexity of time series associated to dynamical systems inferred from independent component analysis," Phys. Rev. E **72**, 046712-1–14 (2005).

[18]M. C. Bishop, *Neural Networks for Pattern Recognition* (Clarendon, Oxford, 1995).

[19]F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence* (Springer-Verlag, Warwick, 1980).

[20]H. D. I. Abarbanel, *Analysis of Observed Chaotic Data* (Springer, New York, 1995).

[21]A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," Phys. Rev. A **33**, 1134–1139 (1986).

[22]M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase space-reconstruction using a geometrical construction," Phys. Rev. A **45**, 3403–3411 (1992)

[23]P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," Physica D **9**, 189–208 (1983).

[24]A. A. Andronov, A. A. Vitt, and S. E. Khaikin, *Theory of Oscillators* Dover, New York, (1966).

[25]A. Maurel, P. Ern, B. J. A. Zielinska, and J. E. Wesfreid, "Experimental study of self-oscillations in a cinfined jet," Phys. Rev. E **54**, 3643–3651 (1996).

[26]H. Hirschberg, J. Kergomard, and G. Weinreich, *Mechanics of Musical Instruments*, CISM Lecture Series (Springer, Berlin, 1996).

[27]Source link: Reiner Janke, http://members.aol.com/ReinerJank/emden.htm. Last viewed 16 July 2007.

Characterization of ultrasound-induced fracture of polymer-shelled ultrasonic contrast agents by correlation analysis

Claudio Pecorari[a)] and Dmitry Grishenkov
Marcus Wallenberg Laboratory, Royal Institute of Technology, Stockholm, Sweden

Beyond a characteristic value of the negative peak pressure, ultrasound fracture the shell of ultrasonic contrast agents (UCAs). Existing criteria for ascertaining this threshold value exploit the dependence of the amplitude of the UCA acoustic response on the incident pressure. However, under the common experimental conditions used in this work, these criteria appear to be unreliable when they are applied to UCAs that are stabilized by a thick polymeric shell. An alternative criterion for determining the onset of shell fracture is introduced here, which uses variations of the shape of the acoustic time-domain response of an UCA suspension. Experimental evidence is presented that links the changes of the cross-correlation coefficient between consecutive time-domain signals to the fracture of the shells, and consequent release of air microbubbles. In principle, this criterion may be used to characterize similar properties of other types of particles that cannot undergo inertial cavitation. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2769618]

I. INTRODUCTION

Ultrasonic contrast agents (UCAs), i.e., gas bubbles stabilized by a solid shell with a diameter of the order of 1 μm, continue to be the subject of considerable research. Their potential role as carriers of drugs to be delivered locally by means of high-power ultrasonic waves is among the applications of highest interest. UCAs with a polymer shell are of particular interest in this context for their considerable chemical versatility and stability.[1–3] In particular, it has been shown that poly (vinyl alcohol) (PVA) can be used to produce microbubble suspensions with a shelf life extending as long as several months.[2] Furthermore, the outer surface of these microbubbles is able to host macromolecules of pharmacological relevance.[3]

Crucial to the effective use of UCAs as drug carriers is the knowledge of the pressure threshold value, P_{thr}, above which the shell fractures. In fact, this quantity determines the upper limit of the pressure range within which the UCAs can be used for contrast-enhancing purposes, although certain imaging techniques exploit the acoustic emission following shell fracture.[4] Above it, high-intensity ultrasonic waves can control localized administration of a drug. To determine P_{thr} of UCA suspensions, purely acoustical methods utilize the peak-to-peak amplitude, A_{pk-pk}, and/or the spectral components, which are not higher harmonics of the fundamental excitation, of the UCA's acoustic response.[5,6] The main assumption behind these methods is that inertial cavitation of UCAs fractures their shells. Inertial cavitation is the violent collapse of a bubble, which follows its expansion beyond a threshold value of its radius. The collapse generates waves with an amplitude considerably larger than that produced by elastic scattering, and with spectral features typical of broadband noise.

Recent optical observations of polymer-shelled UCAs by Bouakaz *et al.*[7] have shown that, during insonation, the shell's radius, R, displays mostly negative variations (compression-only behavior). With a similar technique, Bloch *et al.*[8] have confirmed that the greater stiffness of polymer shells considerably reduces the oscillation amplitude of this novel type of UCAs. Marmottant *et al.*[9] were able to simulate this dynamics by assuming that the shell's effective surface tension is (i) zero for $R < R_{buckling}$, (ii) proportional to R^2, for $R_{buckling} < R < R_{break-up}$, and (iii) equal to the surface tension of the hosting liquid for $R > R_{break-up}$. The symbol $R_{buckling}$ is the radius at which the shell buckles, and $R_{break-up}$ that at which it breaks up. Of relevance for the present work, polymer-shelled UCAs have been observed not to fragment after cracking, retaining their original shape but losing their scattering power if the cracked shell is filled in with liquid.[6]

In this communication, the results of an experimental investigation on the acoustic response of UCAs with a PVA shell are reported, showing that amplitude-based criteria do not offer a viable tool to determine the value of the peak negative pressure, P_{pk}^{-}, at which the shells begin to fracture. An alternative criterion is introduced, which exploits the change of shape of the time-domain acoustic signals acquired upon consecutive insonification of the UCA sample. This criterion uses the maximum value of the cross-correlation coefficient of these signals as a parameter sensitive to the onset of ultrasound-induced shell fracture. Together with a critical interpretation of these results, the link between this parameter and shell fracture is further supported by experiments that also demonstrate the release of air microbubbles following the fracture of the shells.

a)Electronic mail: pecorari@kth.se

II. MATERIALS

In this work, PVA microbubbles with an average diameter and shell thickness equal to 2.7 ± 0.5 and 0.5 ± 0.3 μm, respectively, were used (Cavalieri *et al.*, private communication). The suspension of UCAs was confined within a cylindrical container with a diameter of about 10 mm, and a useful volume of 4 ml. The container wall consisted of a plastic sheet sufficiently thin not to affect the transmission of ultrasonic waves significantly. All the measurements were carried out in de-ionized water at room temperature.

The concentration of microbubbles in the suspensions used in this work was about 1×10^7 ml^{-1}. The suspension was slowly stirred between measurements. The volume in the focal region of the incident pulse that was monitored by the 5 MHz transducer (see Sec. III) was estimated to be of the order of 3 mm^3, and, thus, it contained a few thousand microbubbles.

III. EXPERIMENTAL SETUP AND MEASUREMENT PROCEDURES

Three setups were employed in this investigation. The first one used a focused transducer operating at a nominal frequency of 2.2 MHz, with a diameter of $\frac{1}{2}$ in., and a focal length of 50 mm (Krautkramer, Gamma Series) as emitter, and a 5 MHz, $\frac{1}{2}$-in.-diam, focused transducer with a focal length of 55 mm (Panametrics, V309) as receiver. The incident tone-burst contained six cycles and was emitted at a PRF=500 Hz, unless specified otherwise. The signal driving the transmitting transducer was generated by a high-power, tone-burst pulser-receiver (SNAP Mark IV, Ritec Inc.). The transducers' axes were orthogonal to each other and to the axis of the cylindrical container confining the UCA suspension.

The second setup differed from the first one in that an arbitrary wave form generator (AWG CompuGen 4300, Gage Applied Technologies) in combination with a gated rf amplifier (GA2500, Ritec Inc.) were used in place of the SNAP system. The AWG generated a sequence of three signals—a single sinusoidal cycle, a ten-cycle burst, and a second single cycle—at a distance of 25 μs from each other. Although amplified to the same extent, the amplitude of the single-cycle signals was slightly lower than the maximum amplitude of the ten-cycle burst. The gated rf amplifier allowed for negative peak pressure values, P_{pk}^-, as large as 2 MPa to be generated by the incident waves. In this configuration, the oscilloscope, the function generator, and the gated amplifier were synchronized by a delay generator (565-4CG Berkeley Nucleonics Corporation).

In the third setup, the 5 MHz transducer acted both as emitter and as receiver and was excited by a single cycle signal generated by the SNAP tone-burst generator. The short signals produced in this way reached a maximum peak negative pressure value not exceeding 1 MPa, which was not sufficient to cause the fracture of the shells. The SNAP system was triggered by the delay generator every 500 μs. The 2.2 MHz focused transducer was fed a signal of ten sinusoidal cycles, with a pulse repetition frequency of 100 ms. As in the previous configuration, the signal was amplified by
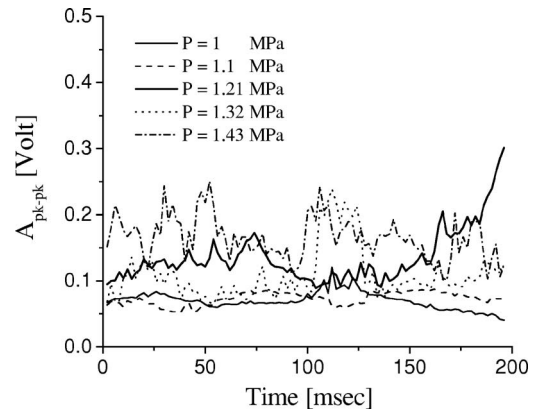

FIG. 1. Peak-to-peak amplitude, $A_{pk\text{-}pk}$, versus time for increasing values of peak negative pressure, P_{pk}^-. The results marked by the thick blue line have been obtained for $P_{pk}^-=1.21$ MPa. Additional system parameters are: pulse repletion frequency, PRF=500 Hz, number of cycles, $N=6$, concentration $=1\times10^7$ ml^{-1}.

the gated amplifier GA2500 to a peak negative pressure of 2 MPa, which was sufficiently high to cause the fracture of the shells and the consequent release of air bubbles in the suspension, as demonstrated by the experiments presented next. The high amplitude signal at 2.2 MHz was delayed by 2.5 ms with respect to the first signal from the 5 MHz transducer. This was done so that the state of the suspension prior to and after the irradiation by the high amplitude burst could be characterized and compared. The delay generator controlled the synchronization of all the above-mentioned components. The time-domain signals were acquired via a 14 bit wave form Digitizer (CompuScope 14200, Gage Applied Technologies) and digitized at a rate of 80 MS/s.

The peak negative pressure, P_{pk}^-, at the focal point of the incident burst was measured by a 75-μm-diam PVdF needle hydrophone (Precision Acoustics) with the cylindrical container removed from the setup. Using this setup, a test was carried out to determine the energy loss affecting the incident beam when it was transmitted through a plastic sheet identical to that used to build the cylindrical container. No appreciable loss of energy was observed.

IV. RESULTS

The large shell thickness of the microbubbles of interest here, together with the shell's elastic properties, frustrates the expansion of the microbubbles which would precede the inertial cavitation and consequent fracture of thin-shelled UCAs and fragmentation of free gas bubbles. For this reason, amplitude-based criteria, which exploit the occurrence of inertial cavitation to determine the pressure at which the shells fracture, can no longer be relied upon to determine the P_{thr} of polymer-shelled UCAs. This point is illustrated by the results of Figs. 1 and 2, which were obtained using the first experimental setup described in Sec. III. In Fig. 1, the time dependence of the peak-to-peak amplitude, $A_{pk\text{-}pk}$, of hundred signals acquired every 2 ms from each other does not clearly define a baseline above which isolated spikes can be identified as cavitation events causing the shell fracture.[6] Figure 2 shows the power spectra which have been averaged over 300 time-domain wave forms that were acquired upon insonifi-

C. Pecorari and D. Grishenkov: Correlation analysis of contrast agents' fracture
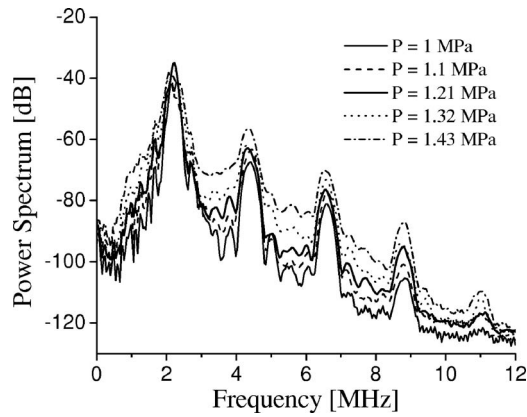
FIG. 2. Average power spectra of the scattered time-domain signals used in Fig. 1. The results marked by the thick solid line have been obtained for $P^-_{pk}=1.21$ MPa. All the system parameters are those given in Fig. 1.

cation of the suspension at the same pressure values reported in Fig. 1. In Fig. 2, the components that fall between higher harmonics are seen to increase with the P^-_{pk} without displaying the expected discontinuity at the onset of the shell fracture. However, it is worth noting in Fig. 1 that just above 1.2 MPa, which is later shown to correspond to the onset of shell fracture, the curves show a distinct change of character, becoming increasingly jagged with increasing pressure. Similarly, in Fig. 2 the increase of not more than 10 dB of the second and third harmonics is accompanied by a variation of the background that is well above 20 dB for an increase of about 400 kPa in the P^-_{pk}. Note that the amplitude
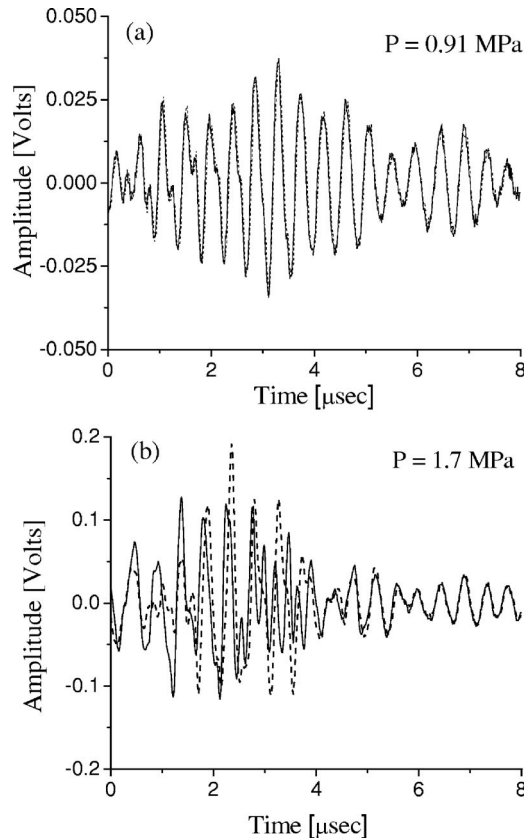

FIG. 3. Examples of time-domain signals due to an incident wave with (a) $P^-_{pk}=0.91$ MPa, and (b) $P^-_{pk}=1.71$ MPa. Other parameters are as in Fig. 1.

of the fundamental nearly does not change. This behavior may be explained at least in part by the nonlinearity of the medium which favors the transfer of the input energy to higher harmonics.

Ultrasonic cracking of polymer shells can also be examined by acoustic means considering the variation of the UCA's response to two consecutive tone bursts. Figure 3 illustrates two examples of time-domain signals acquired 2 ms apart from each other at pressure values below [$P^-_{pk}=0.91$ MPa in Fig. 3(a)] and above [$P^-_{pk}=1.7$ MPa in Fig. 3(b)] the pressure threshold, $P_{thr}=1.21$ MPa. The signals in Fig. 3(a) are identical, reflecting the fact that the configuration of the suspension has not changed during the time elapsed between the arrivals of consecutive bursts, while the opposite is shown in Fig. 3(b).

The similarity of two time-domain signals, $f(t|T)$ and $f(t|T+\Delta T)$, can be measured by the maximum absolute value of the normalized correlation coefficient, $\rho(\tau|T)$, defined by

$$
\max|\rho(\tau|T)|
$$
$$
= \max \left| \frac{\int f(t|T)f(t-\tau|T+\Delta T)dt}{\left\{ \left[\int f^2(t|T)dt \right] \left[\int f^2(t|T+\Delta T)dt \right] \right\}^{1/2}} \right|,
$$

$$(1)$$

where τ is the time lag, T and $T+\Delta T$ are the instants when the consecutive excitations reach the UCAs, and where ΔT $=PRF^{-1}$, PRF being the pulse repetition frequency at which the incident burst is generated. Signals for which Eq. (1) yields a value higher than 0.95 are generated by a suspension that remains largely identical to itself in time. As further discussed in connection with the results presented in Fig. 6, values which are lower than this threshold indicate that the configuration of the system has changed. In Fig. 4, the average value, $\langle\max|\rho(\tau|T)|\rangle$, which was obtained using 99 couples of consecutive signals recorded every 2 ms from T $=0$ ms to $T=198$ ms, begins to decrease steadily below 0.95 as P^-_{pk} increases above 1.2 MPa. The data points in Fig. 4 are average values over three series of measurements, one of
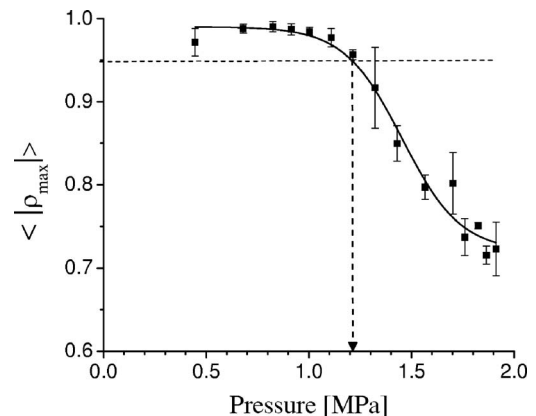

FIG. 4. Average maximum correlation coefficient as function of P^-_{pk}. Other parameters are as in Fig. 1.
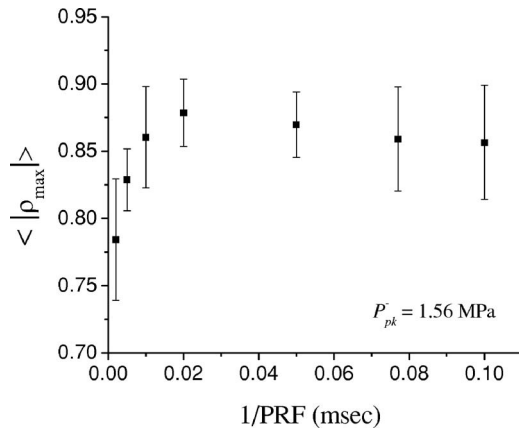
FIG. 5. Average maximum correlation coefficient as function of the inverse of the pulse repetition frequency, PRF. The peak negative pressure was $P_{pk}^- = 1.56$ MPa.
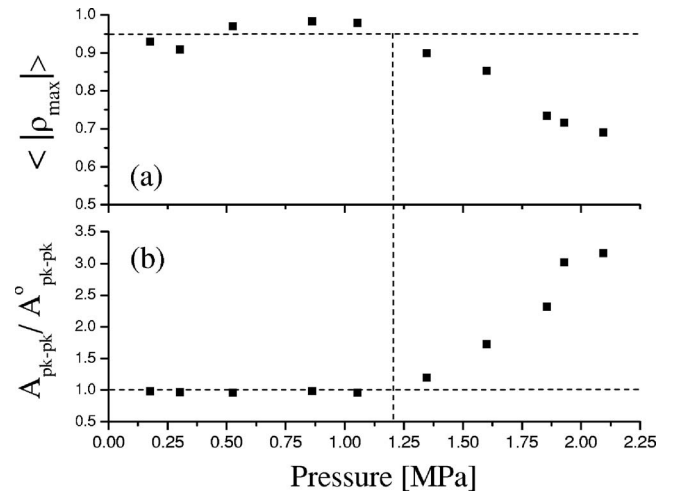

FIG. 6. (a) As in Fig. 4; (b) ratio of the peak-to-peak amplitude of the signals generated by the incident single-cycle pulses insonating the suspension after, $A_{pk\text{-}pk}$, and prior, $A_{pk\text{-}pk}^o$, to the arrival of the ten-cycle, high-amplitude burst.

which has been utilized to obtain the data reported in Fig. 1. The initial increase of $\langle \max|\rho(\tau|T)| \rangle$ at low pressure values has been systematically observed and reflects the contribution of the incoherent, and, thus, uncorrelated noise. This contribution tends to decrease as P_{pk}^- increases. The value of the plateau that the data seem to approach for increasing P_{pk}^- is determined, among other factors, by the coherent component of the scattered signal due to the plastic walls of the containers. The solid line is shown to guide the eye, and has been obtained by best-fitting the data with a sigmoidal curve. Algorithms similar to that given in Eq. (1) have been suggested by Frinking *et al.*[10] for imaging purposes, and used by other authors to monitor changes in the configuration of strongly scattering media.[11–13] Finally, Chomas *et al.*[14] used correlation analysis to the extent the temporal evolution of scattered signals were compared qualitatively to each other in order to investigate mechanisms of destruction of individual microbubbles.

The next three sets of results are presented to demonstrate that the changes in the state of the suspension produced by an incident burst at pressure values above P_{thr} correspond to events in which shells are fractured and air microbubbles are released as a consequence.

Figure 5 illustrates the dependence of $\langle \max|\rho(\tau|T)| \rangle$ on the inverse of the PRF, that is to say, the time between consecutive arrivals of the incident bursts. The data in Fig. 5 represent average values over seven series of measurements. They were obtained using the same configuration employed in the previous experiments and a value of $P_{pk}^- = 1.56$ MPa. The relevant feature of Fig. 5 is the increase of $\langle \max|\rho(\tau|T)| \rangle$ from about 0.78 to a plateau around 0.85 in approximately 10 ms. This time compares well with the dissolution time of air microbubbles released by the fracture of shelled UCAs as measured by Frinking *et al.*[15] The release of air microbubbles explains the increase of $\langle \max|\rho(\tau|T)| \rangle$ with the inverse of the PRF since such highly scattering bodies affect the cross correlation of the signals only as long as the time separation between the latter is shorter than the dissolution time of the air microbubbles. Thus, for PRF higher than

100 Hz, on average, no major contribution of the scattered wave is to be expected from air microbubbles released by the previous burst.

That air microbubbles are released by the UCAs when they are insonified by a burst above the pressure threshold, P_{thr}, is proven by the next two sets of results. Those in Fig. 6 were obtained using the second experimental setup previously described. The choice of using a ten-cycle burst as a second signal was made to increase the impact of the main burst on the configuration of the suspension without exceeding the pressure range used in Fig. 5. The upper part of Fig. 6 was obtained by cross correlating the consecutive time-domain signals produced by the ten-cycle bursts generated every 2 ms. In other words, the procedure that led to Fig. 5 was used in this case as well. The lower part of Fig. 6, on the other hand, reports the ratio between the peak-to-peak amplitude of the second and the first single-cycle signal. This ratio measures the relative variation of the echogeneicity of the suspension caused by the ten-cycle burst. The comparison between the two parts of Fig. 6 shows that the scattering power of the suspension starts to increase at the same pressure at which $\langle \max|\rho(\tau|T)| \rangle$ descends below the threshold value of 0.95. The increase of the scattering power of the suspension is a consequence of the appearance of strong scatterers in the inspected volume. Tests were also carried out using only the two single-cycle pulses to determine whether they had any effect on the integrity of the shells. The results showed that these signals left the microbubbles intact. Note that the results of Fig. 6 provide the physical support for the choice of 0.95 as the threshold value, which indicates the onset of shell fracture introduced in the comments accompanying Fig. 4.

The nature of the scatterers that enhance the echogeneicity of the suspension after the arrival of the high amplitude burst can be recognized by the results of the next figure. Figure 7 shows the time dependence of the peak-to-peak amplitude averaged over nine data sets of the time-domain signals scattered by the suspension. Some of these signal disap-
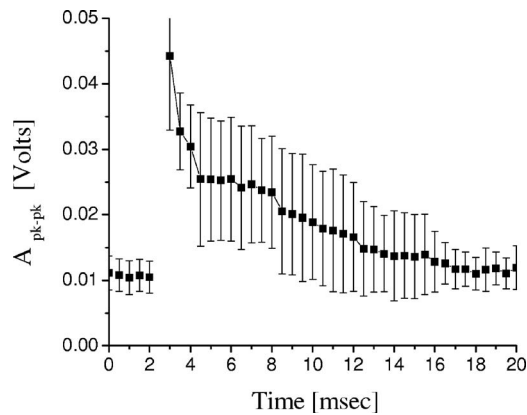
FIG. 7. Peak-to-peak amplitude of the average signal scattered air bubbles released by the polymer-shelled microbubbles upon insonation by an incident burst of ten cycles at a maximum peak negative pressure of 2 MPa.

peared after only 4 ms, others lasted up to 15 ms, which is consistent with the expectation that the scatterers are air microbubbles released by the fractured shell with dimensions distributed within a certain range. These signals were obtained by insonifying the suspension with the 5 MHz focused transducer every 500 μs (PRF=2 kHz) and using the third experimental setup. The arrival of the ten-cycle burst occurred 2.5 ms after that of the first single-cycle, 5 MHz signal. Therefore the data prior to that instant are snapshots of the state of the intact suspension. After the arrival of the high-amplitude burst, the echogeneicity of the suspension suddenly increases, as already shown in Fig. 6, decaying soon after with a characteristic time constant of 9.5 ms. The characteristic time is defined as the time it takes the signal amplitude to decrease to e^{-1} times its initial value. This finding provides further support to the interpretation of the results of Fig. 4.

Reconsidering the results of Figs. 1 and 2, the jagged appearance of the peak-to-peak amplitude curves in Fig. 1 for values of P_{pk}^{-} above P_{thr}, as well as the differential increase of the background with respect to the harmonics in the average power spectra of Fig. 2 can now be naturally accounted for as a result of air microbubbles having been released in the suspension.

Two alternative mechanisms might also be considered for their potential role in the observed loss of correlation. The first one accounts for the microbubbles' buoyancy, while the second is the radiation force induced by the incident high-amplitude wave. The first mechanism can be excluded on the basis of the high cross-correlation measured at low pressure (see Fig. 4). If buoyancy played any significant role, its effect would appear even at low pressure values. This is not the case. The effect of radiation pressure on the motion of UCAs has been investigated both theoretically and experimentally by several authors.[16–19] Their work leads to the conclusion that radiation force produces negligible results under the experimental conditions employed in this work.

V. SUMMARY

The investigation presented in this communication has demonstrated that amplitude-based criteria cannot be relied upon to determine the onset of the fracture of thick, polymer shelled UCAs under the experimental conditions adopted in this work. An alternative criterion based on the cross-correlation analysis of time-domain signals generated by consecutive bursts has been shown to provide a reliable tool for this purpose. The loss of correlation between consecutive signals has been demonstrated experimentally to be caused by the fracture of the shell, which is accompanied by the release of air microbubbles.

A final comment concerns the possible use of the present criterion to determine P_{thr} of systems of particles other than gas-filled microbubbles, which may scatter ultrasonic waves, and also fracture without releasing strong scatterers. An example of such systems is provided by suspensions of micro-capsules with a liquid core.[20] Since the loss of cross correlation between time-domain signals measures the change of the suspension configuration, it is not unreasonable to expect that this criterion may be employed to detect the shell fracture in such systems.

[1]G. Paradossi, F. Cavalieri, E. Chiessi, V. Ponassi, and V. Martorana, "Tailoring of physical and chemical properties of macro- and microhydrogels based on telechelic PVA," Biomacromolecules **3**, 1255–1262 (2002).

[2]F. Cavalieri, A. El Hamassi, E. Chiessi, and G. Paradossi, "Stable polymeric microballoons as multifunctional device for biomedical uses: Synthesis and characterization," Langmuir **21**, 8758–8764 (2005).

[3]F. Cavalieri, A. El Hamassi, E. Chiessi, G. Paradossi, R. Villa, and N. Zaffaroni, "Tethering functional ligands onto shell of ultrasound active polymeric microbubbles," Biomacromolecules **7**, 604–611 (2006).

[4]C. J. Harvey, J. M. Pilcher, R. J. Eckersley, M. J. K. Blomley, and D. O. Cosgrove, "Advances in ultrasound," Clin. Radiol. **57**, 157–177 (2002).

[5]S. L. Poliachik, W. L. Chandler, P. D. Mourad, M. R. Bailey, S. Bloch, R. O. Cleveland, P. Kaczkowski, G. Keilman, T. Porter, and L. A. Crum, "Effect of high-intensity focused ultrasound on whole blood with and without microbubble contrast agent," Ultrasound Med. Biol. **25**, 991–998 (1999).

[6]W.-S. Chen, T. J. Matula, A. A. ′Brayman, and L. A. Crum, "A comparison of the fragmentation thresholds and inertial cavitation doses of different ultrasound contrast agents," J. Acoust. Soc. Am. **113**, 643–651 (2003).

[7]A. Bouakaz, M. Versluis, and N. de Jong, "High-speed optical observations of contrast agent destruction," Ultrasound Med. Biol. **31**, 391–399 (2005).

[8]S. H. Bloch, M. Wan, P. A. Dayton, and K. W. Ferrara, "Optical observation of lipid- and polymer-shelled ultrasound microbubble contrast agents," Appl. Phys. Lett. **84**, 631–633 (2004).

[9]P. Marmottant, S. van der Meer, M. Emmer, M. Versluis, N. de Jong, S. Hilgenfeld, and D. Lohse, "A model for large amplitude oscillations of coated bubbles accounting for buckling and rupture," J. Acoust. Soc. Am. **118**, 3499–3505 (2005).

[10]P. J. A. Frinking, E. I. Cespedes, J. Kirkhorn, H. G. Torp, and N. de Jong, "A new ultrasound contrast imaging approach based on the combination of multiple imaging pulses and a separate release burst," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 643–651 (2001).

[11]O. I. Lobkis and R. L. Weaver, "Coda-wave interferometry in finite solids: Recovery of P-to-S conversion rates in an elastodynamic billiard," Phys.

Rev. Lett. **90**, 354302 (2003).

[12]M. L. Cowan, I. P. Jones, J. H. Page, and D. A. Weitz, "Diffusing acoustic wave spectroscopy," Phys. Rev. E **65**, 066601 (2002).

[13]R. Snieder, "The theory of coda wave interferometry," Pure Appl. Geophys. **163**, 455–473 (2006).

[14]J. E. Chomas, P. Dayton, J. Allen, K. Morgan, and K. W. Ferrara, "Mechanisms of contrast agent destruction," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 232–248 (2001).

[15]P. J. A. Frinking, N. de Jong, and E. I. Cespedes, "Scattering properties of encapsulated gas bubbles at high ultrasound pressures," J. Acoust. Soc. Am. **105**, 1989–1996 (1999).

[16]P. Tortoli, M. Corsi, M. Arditi, and P. Frinking, "Different effects of microbubble destruction and translation in Doppler measurements," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 1183–1188 (2005).

[17]W.-S. Chen, T. J. Matula, and L. A. Crum, "The disappearance of ultrasound contrast bubbles: Observation of bubble dissolution and cavitation nucleation," Ultrasound Med. Biol. **28**, 793–803 (2002).

[18]S. Zhao, M. Borden, S. H. Bloch, D. Kruse, K. W. Ferrara, and P. A. Dayton, "Radiation-force assisted targeting facilitates ultrasonic molecular imaging," Mol. Imaging **3**, 135–148 (2004).

[19]P. A. Dayton, J. S. Allen, and K. W. Ferrara, "The magnitude of radiation force on ultrasound contrast agents," J. Acoust. Soc. Am. **112**, 2183–2192 (2002).

[20]D. G. Shchukin, D. A. Gorin, and H. Möhwald, "Ultrasonically induced opening of polyelectrolyte microcontainers," Langmuir **22**, 7400–7404 (2006).

Measuring fish abundance in a weir trap using an acoustical-optical platform

Jennifer L. Miksis-Olds[a] and Kevin D. E. Stokesbury
University of Massachusetts Dartmouth, School for Marine Science and Technology, 838 South Rodney French Boulevard, New Bedford, Massachusetts 02744

Data recorded by a bottom mounted survey platform deployed within the opening of a fishing weir were used to calculate species specific abundance estimates for comparison to the weir catch. Abundance estimates were calculated from the combination of sonar and video information recorded by the Acoustical-Optical Platform (AOP). Echo counting was used to detect individual moving targets in the sonar images with the application of a background removal technique utilizing a moving average filter. Video images provided species identification of acoustic targets. Video images and differences in target strength distributions reflected a change in dominant species from each deployment which was confirmed with the weir catch. The algorithm used to calculate AOP abundance estimates was most accurate in predicting abundance for species comprising at least 13% of the overall catch by weight. Close agreement between the species specific AOP estimates and absolute abundances of each species suggests that the combination of acoustic and video data is a powerful combination for accurately identifying fish species and predicting abundance. © 2007 Acoustical Society of America. [DOI: 10.1121/1.2775426]

PACS number(s): 43.80.Ev, 43.30.Sf [WWA] Pages: 2431–2438

I. INTRODUCTION

The Acoustical-Optical Platform (AOP) was designed to meet a management need for more accurate stock assessments of pelagic and groundfish species. The objective of this study was to evaluate the accuracy of AOP abundance estimates by comparing species specific abundance estimates obtained from AOP data to the absolute species specific abundances of a fishing weir. The weir trap provided the optimal setting for a performance evaluation of the AOP because of the finite sampling volume and absolute catch count for comparison to the AOP abundance estimates. Similar evaluation is intractable offshore due to the boundless sampling volume and inability to define an absolute catch. The fishing weir also provided a unique opportunity to evaluate the AOP performance in a setting where multiple species were encountered. Although the shallow coastal species in the weir catch differed from the species most likely encountered offshore, the sampling methodology used in the present study is directly applicable to offshore sampling of pelagic and groundfish species.

II. METHODS

The AOP, recently developed at the University of Massachusetts Dartmouth, School for Marine Science and Technology, is a bottom mounted survey platform that provides data from which species composition can be identified, fish length can be calculated, and fish abundance can be estimated (Gröger *et al.*, 2004; Miksis-Olds and Stokesbury, 2006). The platform was deployed at the mouth of a fishing weir to monitor fish swimming into the trap. The number of fish estimated to have passed the AOP was then compared with the number of fish caught in the trap.

A. Acoustical-optical platform

The modular, three-tiered galvanized steel AOP frame was equipped with two low light DeepSea cameras, four laser references, two 200 kHz phased array sonar transducers, solid state light, battery package, and environment sensors including light, temperature, and depth sensors (Fig. 1). The base dimensions were approximately $2 \text{ m} \times 2$ m, and the pyramid stood 1.5 m tall. Light levels below 41 Lux resulted in poor video quality, so the solid state light was triggered on below this threshold to maintain high video image quality. The sonar system, originally designed for obstacle avoidance, was commercially available from Interphase and had been modified to record and store target echo levels (Bernard *et al.*, 2006). One transducer was mounted in the horizontal direction to image the area directly above the seafloor. The second transducer was oriented vertically to sample the water column above the platform. Both transducers were calibrated using the standard target calibration technique (Foote and MacLennan, 1984). The AOP cameras were low light, black and white cameras from DeepSea Power and Light (Multi-Seacam 1060). The cameras were positioned so that the volume of water sampled by the horizontally oriented camera overlapped with the horizontal transducer, whereas the vertical camera was positioned to overlap with the vertical transducer. Calibration performed in an acoustic test tank showed that 98% of the video image overlapped with the volume sampled by the sonar at 6 m when the camera was angled off the vertical toward the corresponding transducer by $5° – 15°$.
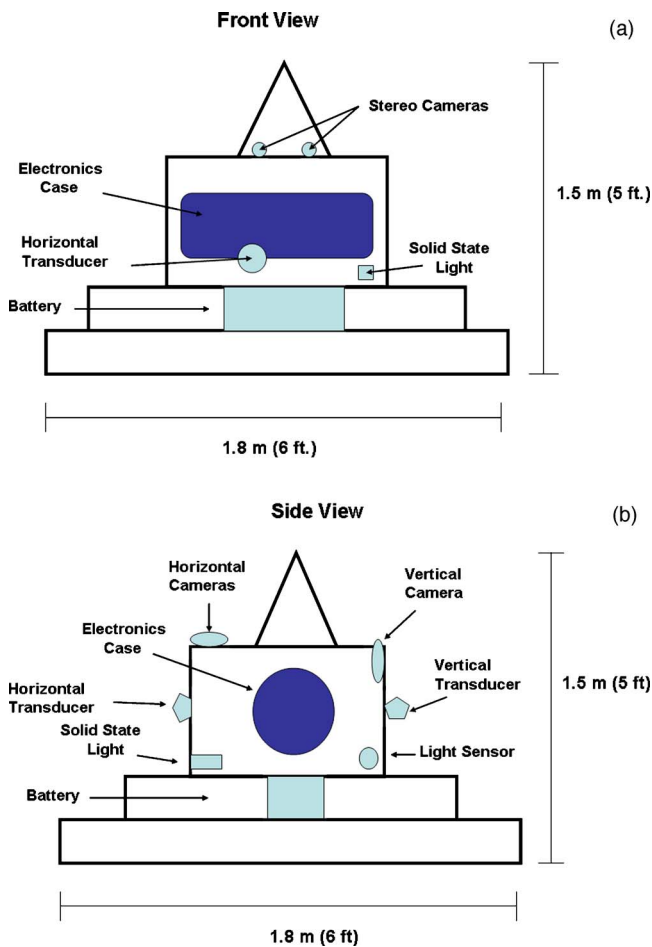
FIG. 1. (Color online) Schematic of the Bailey weir layout and AOP position. The view in (a) is from the top down perspective. The view in (b) is a cross-sectional perspective looking down between the leader and wing nets. The patterned area represents the sampled area.

Two lasers were also positioned within the field of view of each camera to provide length estimates of fish that simultaneously intersected both lasers.

B. Weir deployments

The AOP was deployed in the mouth of a fishing weir off Sakonnet Point, RI [Fig. 2(a)] three separate times: 30–31 May; 7–9 June; and 13–14 June. The Bailey weir, owned and operated by H.N. Wilcox, was approximately 14 m deep, 20 m wide, and 50 m long measured from the tip of the wing net to the end of the catch box. The 500 m leader net with 19 cm (7.5″) mesh extended toward shore. The wing net was constructed of 10.2 cm (4″) mesh, and the mesh in the catch area was 5.1 cm (2″). Data collection during each deployment began after the morning removal of the previous day's catch and ended after the removal of the surveyed catch. The trap was hauled after 24 h during the first and third deployments. The weir was not hauled at the 24 h mark during the second deployment due to adverse weather conditions, so the second deployment was extended to cover a 48 h time period. The AOP was deployed and retrieved from the seafloor using a winch system on the F/V H.N. Wilcox. After the AOP was placed on the seafloor inside the wing net, divers
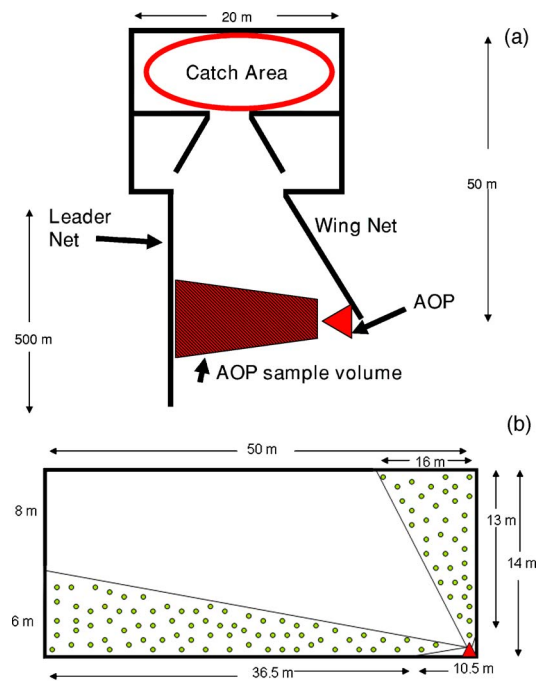

FIG. 2. (Color online) AOP assembly. The front view (a) and side view (b) of the AOP.

oriented the AOP so the sonar and video sensors were aligned to sample a cross-sectional area between the wing and leader nets (Fig. 2).

The sonar range was set at 50 m to cover the full cross-sectional area of the weir mouth (Fig. 2) and the full depth of the water column which oscillated between 12 and 16 m dependent on tidal cycle. The leader net and water surface were clearly visible on the horizontal and vertical sonar images, respectively. This confirmed full acoustic coverage between the AOP sonar and leader net in the horizontal direction and also between the AOP sonar and surface in the vertical direction. The beamwidth of each sonar beam was 12° in the vertical and 11° in the horizontal. Beamforming produced 31 sequential beams at 31 different, but overlapping, angles that combined to cover a full 90° in the horizontal direction in one sonar scan. The AOP sonar sampled approximately 40% (280 m^2) of the cross-sectional area (700 m^2 calculated from a 50 m range and average 14 m depth) between the wing and leader nets, based on beamwidth geometry, height of the transducers from the seafloor, and range [Fig. 2(b)]. The sonar system transmitted ten consecutive pulses during a 1 min sampling period. This was followed by a 15 min "sleep" interval for power conservation. The pulse length was 325 μs. The pulses were spaced approximately 6 s apart. This resulted in the sonar system cycling through a pattern that transmitted ten pulses over a 1 min sampling period and entered a power conservation mode for 15 min to produce an overall 16 min cycle. The backscattered energy received from each individual pulse is referred to as a sonar scan.

A video image was captured from each camera at 30 s intervals to ensure overlap with each sonar cycle. The range and volume sampled by the video was variable due to con-

stantly changing water clarity and light conditions. It was estimated that the maximum range observed by video was 10 m under conditions of greatest water clarity and strongest light levels. The fixed focus cameras had a field of view under water of $57° \times 45° \times 71°$. All data were recorded to an embedded computer and downloaded post-retrieval.

Following each haul, fish were separated by species, and the total weight of each species caught was recorded. For species subject to length or weight restrictions, further discrimination was made between individuals meeting the take criteria and those categorized as discard. Weights and length measurements of commercially valuable species were taken for each discarded individual. For species in which less than 100 individuals were kept, all fish were weighed and measured. Subsampling of approximately 100 individuals per species occurred for species with allowable catch counts greater than 100 individuals. Subsampling procedure adhered to the protocol outlined in the Northeast Fisheries Observer Program (2007) and in Murphy and Willis (1996).

C. Data analysis

Analysis of backscattered energy in each sonar scan was a two-step process: (1) background removal, and (2) target detection. Background removal was accomplished via a moving average filter with a sliding window of 40 scans. This is a common method used in signal processing for removing noise from time series or images (Dunstan, 1993; Bendat and Piersol, 2000). The theory behind background removal with a moving average filter is to separate, or subtract out, stable reoccurring features (the background) from changing foreground features in a scan of interest (Wu and Jeng, 2002). A separate background image was created for each sonar scan analyzed by averaging 40 total scans, 20 prior to the scan of interest and 20 directly following the scan of interest. The average background image created by the moving average filter was composed of stable, reoccurring features such as rocks, shells, and debris on the seafloor and stable characteristics of the water column. Subtracting the background from each sonar scan resulted in the detection of moving targets in the foreground pixels with units of decibels. Moving targets were classified as fish based on the decibel difference between the background image and the viewed scan. Single fish echoes were defined as echoes with lengths less than twice the pulse wave length. All moving targets with target strengths between −20 and −75 dB were classified as potential fish. Potential fish targets were verified visually, and the range and bearing of each target was calculated at the position of the target's maximum backscattered energy with MATLAB processing software.

Video images were viewed manually by an analyst blind to the species composition of each weir catch. Initial inspection separated images into two categories: target absent and target present. Images with targets were then viewed a second time for species identification (Fig. 3). Video images were captured every 30 s, whereas sonar sampling took place for 1 min every 15 min. The video sampling interval (30 s) was approximately 30 times greater than the sonar sampling interval (16 min), so the video identification of any indi-
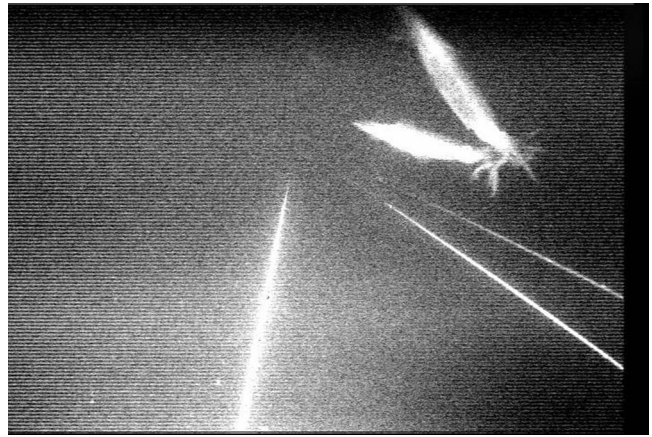

FIG. 3. Video image of squid, *Loligo pealei*, the dominant species in Deployment 1.

vidual within the 16 min sonar interval was extrapolated to all sonar targets within that particular interval. There were no instances where more than one species was identified in video images within the same 16 min interval. This eliminated the need to proportion the acoustically detected targets to account for multiple species detected within a sampling interval for this study. The extrapolation of identified targets to the 16 min interval did create a bias against identification and resulting abundance estimates of unexpected or hard to see targets, and this bias will need to be addressed in data sets that contain video identification of more than one species within a sampling interval.

Species specific abundance estimates were calculated from the number of sonar targets detected by the horizontal and vertical transducers and verified by video within each 16 min interval. An AOP Abundance Estimate (n_{AOP}) was calculated according to Eq. (1). The Corrected Abundance Estimate ($n_{corrected}$) was obtained by multiplying the AOP Abundance Estimate by 2.5 to reflect the 60% of the cross-sectional area not sampled [Eq. (2), Fig. 2]. Equation (3) reflects the simplified combination of Eqs. (1) and (2). This algorithm operated on two major assumptions: (1) all the targets detected in a single sonar scan were different individuals, and (2) individuals were uniformly distributed over the cross-sectional area of the weir opening

$$n_{AOP} = (\bar{x}_x + \bar{x}_y) I [16(m_{xy})], \tag{1}$$

where x_x=horizontal encounter rate (individuals/min), x_y=vertical encounter rate (individuals/min), I=number of intervals a specific species was detected, and m_{xy}=proportion of sonar scans a species was detected in the ten-scan sampling period in either the horizontal or vertical direction,

$$n_{corrected} = 2.5 n_{AOP}, \tag{2}$$

$$n_{corrected} = 40 I (\bar{x}_x + \bar{x}_y)(\bar{m}_x y). \tag{3}$$

The first term in Eq. (1) ($\bar{x}_x + \bar{x}_y$) refers to the average encounter rate. Following a positive species identification within a 16 min sonar interval, the maximum number of sonar targets within a single scan during the 1 min sampling

TABLE I. Summary of weir catches. An asterisk denotes abundance estimates calculated from the mean weight of species specific subsamples averaged across deployments.

Deploy		American Shad	Black Sea Bass	Blue fish	Fluke	Loligo Squid	Scup	Sea Robin	Skates	Striped Bass	Tautog	Torpedo Ray	Fourspot Flounder	Other	Total
1	Total wt (kg)	0.9	13.9		124.5	1225.8	232.7	44.6	159.0	26.5	4.9		5.4	5.8	1846
	Measured?	N	Y		Y	N	Y	N	Y	Y	Y		N		
	Subsample wt (kg)	···	13.9		63.8	···	76.9	···	159.0	26.5	4.9		5.4		
	No. in subsample	1	12		113	···	217	···	4	41	3		24		
	Mean wt (kg)	0.9	1.2		0.5	···	0.4	···	39.7	0.6	1.6		0.2		
	Proportion (%)	0.05	0.8		6.7	66.4	12.6	2.4	8.6	1.4	0.3		0.3	0.3	100
	Abundance (indiv.)	1	12		228*	9000*	559	···	4	41	3		24	13	9885
2	Total wt (kg)		10.4	24.1	52.0	4.5	8.2	5.2	17.9	459.9	2.0	99.9		22.7	684
	Measured?		Y	N	Y	Y	Y	N	N	Y	Y	N			
	Subsample wt (kg)		10.4	···	52.0	4.5	8.2	···	···	459.9	2.0	99.9			
	No. in subsample		7	···	76	38	13	···	···	247	1	4			
	Mean wt (kg)		1.5	···	0.6	0.1	0.6	···	···	7.6	2.0	25.0			
	Proportion (%)		1.5	3.5	7.6	0.7	1.2	0.8	2.6	67.2	0.3	14.6		3.2	100
	Abundance (indiv.)		7	···	76	38	13	···	···	247	1	4			386
3	Total wt (kg)	4.1	17.7	77.2	23.6	1.6	1.6	18.8	9.3	83.8	1.4	326.9	0.5	10.6	577
	Measured?	Y	Y	Y	Y	Y	Y	N	N	Y	Y	N	N	Y	
	Subsample wt (kg)	4.1	17.7	77.2	23.6	1.6	1.6	18.8	9.3	83.8	1.4	326.9	0.5	5.2	
	No. in subsample	7	15	155	40	12	3	···	···	34	1	12	1	1	
	Mean wt (kg)	0.6	1.2	0.5	0.8	0.1	0.5	···	···	7.1	1.4	27.2	0.5	5.2	
	Proportion (%)	0.7	3.1	13.4	4.1	0.3	0.3	3.3	1.6	14.6	0.2	56.7	0.1	1.9	100
	Abundance (indiv.)	7	15	155	40	12	3	···	···	34	1	12	1	4	284

period was used to calculate an encounter rate/minute within a single interval. The maximum number of targets within a single scan during any 16 min interval, as opposed to the sum or average number of targets detected in the interval, was selected as the parameter of estimation to ensure that no target was counted twice. This is a conservative approach used to eliminate the argument that a target seen in one scan could be one of the same targets seen in consecutive scans. Encounter rate was computed separately for the horizontal and vertical transducers, as targets in one view could not be present in the other view during simultaneous sampling. An overall species specific encounter rate for each transducer orientation was calculated by averaging rates from each interval in which a particular species was present. The rates of the two orientations were then added together.

The final term in Eq. (1) $[16(m_{xy})]$ is the average number of minutes per interval an individual of a particular species was estimated to be present. This term was determined by averaging the proportion of sonar scans in which species specific targets were identified in either the horizontal or vertical direction during each 1 min sampling period. This proportion was then multiplied by 16, the number of minutes in each interval, to determine the overall number of minutes a species was estimated to be present.

III. RESULTS

The overall catch weight and number of individuals was greatest during the first deployment (1846 kg and approximately 9885 individuals) and decreased with each deployment. Details from each catch are shown in Table I. The species composition changed from deployment to deployment. The first deployment was dominated by squid (*Loligo pealei*), whereas the second deployment was dominated by striped bass (*Morone Saxatilis*). According to weight, the third deployment was dominated by 12 torpedo rays (*Torpedo nobiliana*) with an average individual weight of 27 kg. However, according to the number of individuals per species in the catch, the third deployment was dominated by bluefish (*Pomatomus Saltatrix*).

The number of targets detected by the sonar and video sensors reflected the overall pattern of catch weight and number of individuals with decreasing number of targets over the course of the deployments (Table II). Sonar and video detections were positively correlated with overall catch weights. The values of the correlation analysis were 0.82 and 0.99 for sonar data and video data, respectively. However, the small sample size ($n=3$) did not provide adequate power for a linear regression analysis. The combination of sonar and video data successfully identified the dominant species in each catch, with the exception of flatfish (Tables I and III). Species identification from video images ranged from 0.02% to 2.3% over the three deployments (Table II). Sonar target detection ranged from 48% to 64% over the course of the three deployments (Table II). The horizontal sonar and cameras consistently revealed a greater percentage of targets than the vertical components, most likely due to the greater volume of water sampled in the horizontal direction.

Each deployment revealed a different signature in terms of diurnal pattern of target detection and distribution of target strengths. Sonar target detection also showed a diurnal pattern with an increase in target detections during the night hours for the first deployment. The average number of targets detected within each interval during the day hours (6:00–18:00) was significantly less than the average number of tar-

TABLE II. Summary of target detections from video and sonar sensors. Vertical sensors are denoted by (V), and horizontal sensors are denoted by (H). Total values computed from both sensor orientations is denoted by (T). The percentages shown in parentheses indicate the proportion of images in reference to the respective sensor orientation or total images. The integral number in parentheses under Species specific image ID represents the number of video images from the combined horizontal and vertical cameras containing each species.

Deployment	Data type	Total samples	Samples with targets	Images with identifiable species	Species specific image ID
1	Video	(H) 2576	(H) 262 (10%)	61 (2.3%)	Squid (45)
30–31 May 2006		(V) 2576	(V) 10 (0.3%)	2 (0.08%)	Scup (18)
		(T) 5152	(T) 272 (0.05%)		
	Sonar	(H) 850	(H) 794 (93%)		
		(V) 850	(V) 332 (39%)		
		(T) 1700	(T) 1126 (64%)		
2	Video	(H) 4698	(H) 130 (3%)	19 (0.4%)	Squid (6)
7–9 June 2006		(V) 4698	(V) 3 (0.06%)	1 (0.02%)	Scup (1)
		(T) 9396	(T) 133 (0.01%)		Striped bass (10)
	Sonar	(H) 1662	(H) 1352 (81%)		Tautog (3)
		(V) 1662	(V) 505 (30%)		
		(T) 3324	(T) 1857 (56%)		
3	Video	(Hs) 2391	(H) 17 (0.7%)	11 (0.5%)	Squid (11)
13–14 June 2006		(V) 2391	(V) 3 (0.1%)	2 (0.08%)	Bluefish (2)
		(T) 4782	(T) 20 (0.004%)		
	Sonar	(H) 827	(H) 485 (59%)		
		(V) 827	(V) 255 (31%)		
		(T) 1654	(T) 740 (48%)		

gets detected within each interval during the night hours (18:00–6:00), as demonstrated by a standard t-test (t-stat= −2.62, p=0.01, d.f.=82). The average number of detections per interval during the day was 61±40 individuals per interval compared to a mean of 82±33 detections at night. The reverse pattern was observed during the second deployment. The average number of targets detected within each interval during the day hours was significantly greater than the aver-age number of targets detected within each interval during the night hours (t-stat=3.23, p=0.001, d.f.=160). The aver-age number of detections per interval during the day was 36±22 individuals per interval compared to a mean of 26±16 detections at night. There was no significant diurnal pattern observed in the third deployment. The average num-ber of sonar detections per interval was 14 for both the day and night.

TABLE III. Summary of species specific abundance estimates between the AOP and weir catch. Columns 3–5 itemize the values used in the AOP abundance calculation algorithm. % error represents the percent difference between the AOP estimate and absolute abundance from the weir catch. Abundance estimates calculated for Deployments 1, 2, and 3 used a −60 dB target strength threshold. Abundance estimates for Deployment 1[*] were calculated without the application of any threshold.

Deployment	Species	No. intervals detected	Encounter rate (indiv/min)	Minutes present	Corrected AOP abundance est.	Catch abundance	% error (%)	Proportion of catch (%)
1	Loligo squid	20	8	16	6426	9000	−29	66
	Scup	4	4	14.4	576	559	3	13
1[*]	Loligo squid	20	12.3	16	9818	9000	9	66
	Scup	4	6.5	14.4	936	559	67	13
2	Loligo squid	6	4.7	12.8	896	38	2258	0.6
	Scup	1	12	16	480	13	3592	1
	Striped Bass	1	8	14.4	288	247	17	65
	Tautog	2	7.5	16	600	1	59900	0.3
3	Loligo squid	2	3	9.6	144	12	1100	0.3
	Bluefish	1	15	4.8	180	155	16	13

FIG. 4. (Color online) Target strength distributions of each deployment.
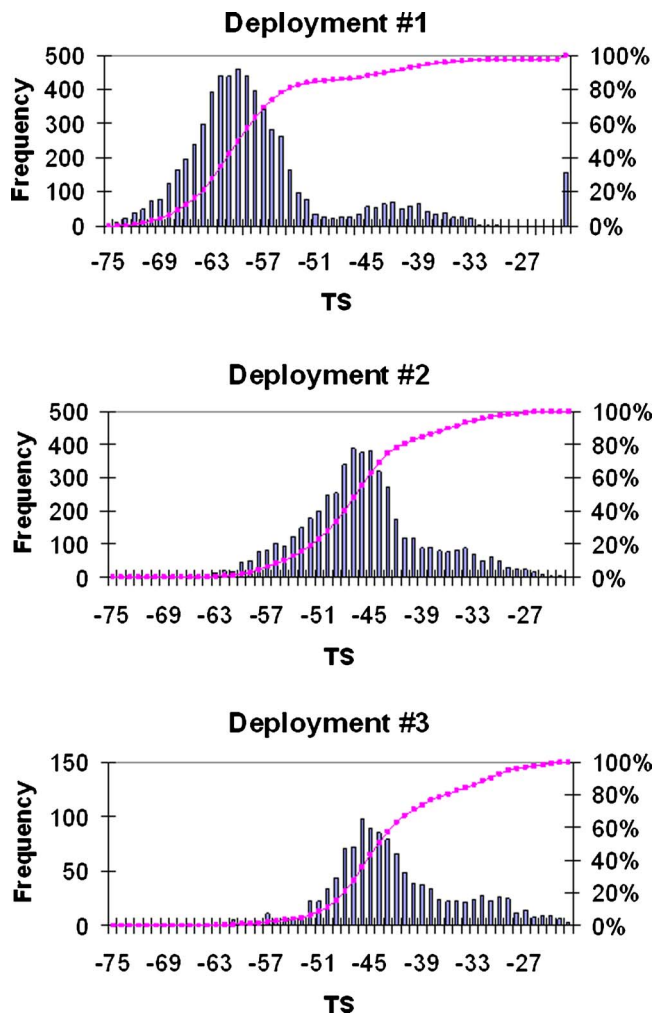

FIG. 5. (Color online) Range distributions for each deployment.

Target strength distribution was also unique to each deployment. The average target strength was −57, −46, and −43 dB for the first, second, and third deployments, respectively. A Lilliefors test for normality showed that each distribution was not normally distributed and suggested multiple modes (Fig. 4). Each target strength distribution was significantly different from every other as indicated by a series of Bonferroni corrected Kolmogorov-Smirnov two-sample tests ($p < 0.01$ for each corrected comparison). Acoustic target range distributions were also constructed for each deployment. There were visible differences between each of the range distributions, but each distribution showed a bimodal pattern with peaks within a few meters of the AOP and approximately 35–45 m from the platform (Fig. 5). The acoustic signatures of the nets themselves were removed from the foreground with the moving average filter so that the resulting distributions reflected only the detected animals. The first peak corresponds to animal detections in close proximity to the wing net, whereas the second peak corresponds to animals near the leader net (Fig. 2). The variability of the second range peak is most likely associated with the movement of the leader net due to tides and currents.

Initial abundance estimates from all deployments were made with the application of a −60 dB target strength threshold; all 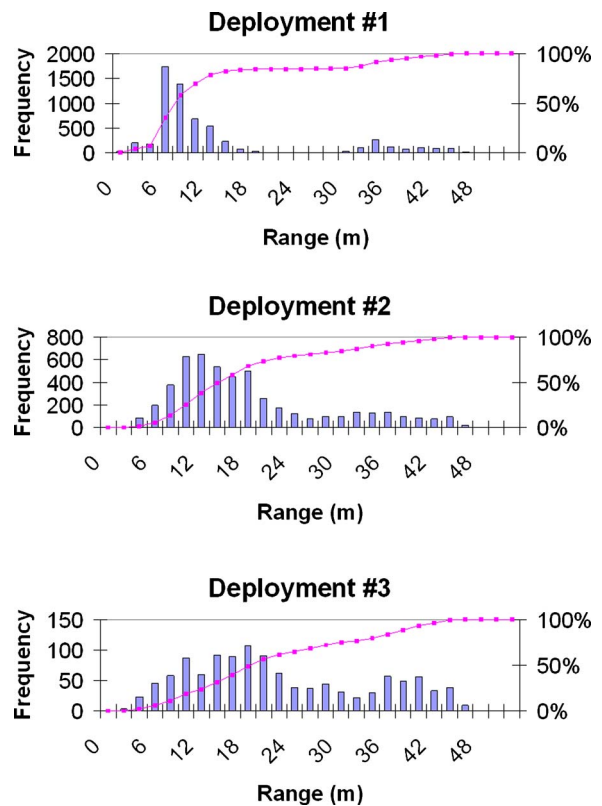targets with a target strength greater than −60 dB were included in the estimate calculation. The −60 dB threshold was selected to coincide with a conservative range of dorsal aspect target strengths for swim-bladdered fish (Benoit-Bird and Au, 2001; Simmonds and MacLennan, 2005); analogous data for lateral and/or ventral aspect target strengths were not available. Abundance estimates using the −60 dB threshold tended to underestimate the catch abundance in Deployment 1 and overestimate the catch abundance in Deployments 2 and 3 (Table III). The abundance estimates in all deployments were sensitive to the number of individuals of each species catch and the proportion of catch that each species comprised (Fig. 6). Estimates were most accurate when more than 50 individuals of each species was caught, resulting in a proportion of at least 13%. Estimates for identified species with target strengths greater than −60 dB and composing at least 13% of the total catch by weight ranged from −29% to 17%. Deployment 1 differed from the other two deployments in that it underestimated the abundance. Based on observation of the target strength distribution, the application of a −60 dB threshold excluded a majority of the targets detected. The dominant species in Deployment 1 was identified by video as squid, which has target strengths ranging from approximately −80 to −55 dB (Jefferts et al., 1987; Benoit-Bird and Au, 2001; Goss et al., 2001). Recalculating the abundance estimates for Deployment 1 with no target strength threshold, in order to incorporate the less reflective squid more accurately, resulted in abundance estimates that were overestimates of 9% and 67% error for squid and scup, respectively (Table III, Deployment 1*).
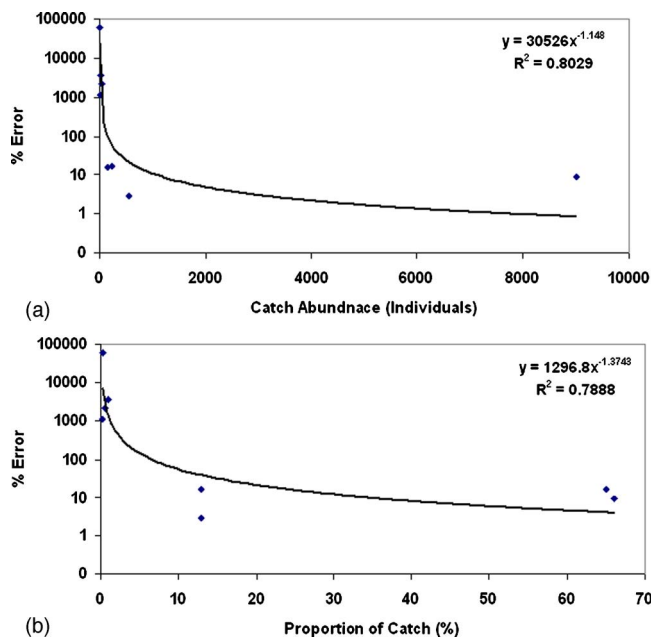
FIG. 6. (Color online) Relationship between percent error of the abundance estimate and the weir catch. Each point represents the species specific abundance estimates presented in Table III. (a) Percent error and weir catch abundance relationship depicted with a best fit power curve. (b) Percent error and proportion of catch relationship depicted with a best fit power curve.

IV. DISCUSSION

The AOP capitalizes on the complementary strengths of acoustical and optical sampling techniques by using acoustics for fish enumeration and optics for species identification and measurements of fish length during multiple day deployments. Simultaneous collection of acoustical and optical data from the same parcel of water also eliminates the need for separate, labor intensive groundtruthing efforts. Monitoring fish from the ocean floor can increase target detection and eliminate or reduce many of the availability biases of trawl and acoustic surveys employing hull mounted echosounders (Aglen, 1996; Simmonds, 2003; Simmonds and MacLennan, 2005). However, the AOP is still subject to biases relating to target echo level changes due to swim bladder size, depth, tilt angle, and shape to abundance estimates (Hazen and Horne, 2003; Misund, 1997; MacLennan, 1990). The video component of the AOP data complements the acoustic data allowing many of these biases to be addressed. Target strength, species, size, and observation angle are quantified when a fish is simultaneously detected by both the sonar and optical sensors of the AOP (Bernard et al., 2006; Miksis-Olds and Stokesbury, 2006). When fish are simultaneously observed on the sonar and the video sensors, data that will relate the fish size and observation angle to target strength are accumulated. This information is vital to calculating the most accurate estimates of fish abundance and biomass (Simmonds and MacLennan, 2005, Hazen and Horne, 2003; Misund, 1997; MacLennan, 1990).

Stationary acoustic sampling has previously been used in rivers and weirs to monitor and assess the passage of a single targeted genus such as Pacific salmon smolt (*Oncorhynchus spp.*) (Nunnallee, 1983; Ransom et al., 1996).

Acoustic survey and net catch estimates correlated with a high degree of accuracy ($r=0.96$) (Ransom et al., 1996), but individual target detection at the species level was not possible. The video component of the AOP provides the necessary data to identify acoustic targets at the species level in an environment where multiple species are encountered.

The combination of video and acoustic data recorded by the AOP at the mouth of a fishing weir permitted the discrimination of targets at the species level. The dominant species identified from the AOP video matched the dominant species in the weir catch for all deployments. Species specific abundance estimates of the weir catch from AOP data were most accurate when the target species composed at least 13% of the total catch by weight. These data also suggested a pattern of positive correlation between the number of video and acoustic target detections and the overall catch weight of each deployment's catch; however, this pattern needs to be confirmed with a larger sample size. Additional information for discriminating between species was obtained from the acoustic analysis. Significant differences in the target strength distributions suggest that the species composition changed in each deployment, which was verified by both the video data and weir catch. Differences in target strength distributions are mostly attributed to differences in target anatomy and physiology (Simmonds and MacLennan, 2005). The acoustic analysis of AOP data also showed a difference in diural pattern between deployments indicative of different species specific behaviors. In the first deployment more targets were observed at night, and the catch was dominated by squid. Squid exhibit a behavior in which they are attracted to light which reflected the pattern observed during the first deployment (Hanlon and Messenger, 1996). The second deployment showed a reverse pattern with more targets detected during the day. Consequently, the squid composition was less than 1% of the total catch.

Deploying the AOP in an environment where the absolute number[1] of fish passing by the platform also revealed a number of AOP shortcomings that need to be addressed before it can accurately be applied to abundance estimates of fish in the open ocean. There was a gross overestimation of abundance for species in which only a few individuals were caught. This was most likely attributed to the same individual(s) being recorded by the video and acoustics during multiple time periods. It was possible to eliminate this bias within intervals by using the maximum number of individuals detected in one video image/sonar scan as the abundance parameter, but it may be impossible to completely eliminate the possibility of counting the same individual in multiple intervals. There is also the possibility of underestimating the abundance if the first assumption of the abundance algorithm is violated. It was assumed that each target represented only one fish, but if two or more fish were not separated by the minimum resolution of the system, they would appear as only one target; thus the estimate would underestimate the true number of fish. In addition, if the AOPs constitute a fish aggregation device (FAD) for particular species, such as squid, this behavior will need to be accounted for in the overall abundance estimates.

In addition to information pertaining to abundance, analysis of the range distribution of acoustically detected targets provided some insight into how the fish were distributed within the mouth of the weir opening. The range distribution of targets was bimodal with peaks corresponding to higher abundance of targets at the wing and leader nets. They were not uniformly distributed throughout the weir opening, which violates the second assumption of the abundance calculation. More accurate abundance estimates can be obtained by accounting for the nonuniform distribution in the future. Monitoring range distributions have the potential to supply further information on associated behavior patterns. For example, an analysis of species specific range distribution with a larger data set will provide one technique for evaluating whether or not the AOP is a FAD for that species. Quantitative data relating fish abundance and distribution within the leader-wing net area may also be useful in designing weirs to more effectively target specific species.

Having identified specific strengths and weakness of the AOP in a relatively controlled environment, the transition to its application in offshore waters can begin. One AOP has a limited sonar and video spatial scope of approximately 2000 and 20 m^2, respectively. However, the AOP is designed to be used in an observation system with multiple platform elements throughout an area, so fish populations can be monitored over greater spatial scales (Gröger *et al.*, 2004; Miksis-Olds and Stokesbury, 2006). Continual advances in video processing will also play a large role in increasing the number of images in which species can be identified and length can be measured. With a larger sample size of clear images, target strength-length relationships can be generated from targets simultaneously imaged by sonar and stereo cameras. For many species, length corresponds well to age enabling age-specific abundance estimates.

ACKNOWLEDGMENTS

[1]The absolute number of fish passing by the AOP and being reflected in the weir catch is subject to the caveat that fish detected by the AOP could have been small enough to escape the weir trap or swim out of the weir mouth before being trapped.

Aglen, A. (**1996**). "Impact of fish distribution and species composition on the relationship between acoustic and swept-area estimates of fish density," ICES J. Mar. Sci. **53**, 501–505.

Bendat, J. S., and Piersol, A. G. (**2000**). *Random Data Analysis and Measurement Procedures* (J Wiley, New York), 594 pp.

Benoit-Bird, K. J., and Au, W. W. L. (**2001**). "Target strength measurements of Hawaiian mesopelagic boundary community animals," J. Acoust. Soc. Am. **110**, 812–819.

Bernard, E., Jakubiak, C. J., Miksis-Olds, J. L., Penvenne, J., and Holliday, D. V. (**2006**). "Calibration of a steered phased-array sonar for use in fish detection," OCEANS '06 MTS-IEEE Conference, Boston, MA, September 18–22.

Dunstan, F. D. J. (**1993**). "Time series analysis," in *Biological Data Analysis*. edited by J. C. Fry (Oxford University Press, New York), pp 243–310.

Northeast Fisheries Observer Program. (**2007**). *Fisheries Observer Program Manual* (U.S. Department of Commerce/NOAA Fisheries Service), http://www.nefsc.noaa.gov/femad/fishsamp/fsb/. Last viewed 8/24/07.

Foote, K. G. and MacLennan, D. N. (**1984**). "Comparison of copper and tungsten carbide calibration spheres," J. Acoust. Soc. Am. **75**, 612–616.

Goss, C., Middleton, D., and Rodhouse, P. (**2001**). "Investigations of squid stocks using acoustic survey methods," Fisheries Research **54**, 111–121.

Gröger, J. P., Truong, T., Rothschild, B. J., and Azadevar, F. (**2004**). "A new approach towards an optimal control of fisheries on Georges Bank: The SMAST groundfish optimization simulation model," Vigo, Spain, September 22–25, ICES CM 2004/ FF 18.

Hanlon, R. T., and Messenger, J. B. (**1996**). *Cephalopod Behaviour* (Cambridge University Press, Cambridge), 232 pp.

Hazen, E. L., and Horne, J. K. (**2003**). "A method for evaluating the effects of biological factors on fish target strength," ICES J. Mar. Sci. **60**, 555–562.

Jefferts, K., Burczynski, J., and Pearcy, W. G. (**1987**). "Acoustical assessment of squid (*Loligo opalescens*) off the central Oregon coast," Can. J. Fish. Aquat. Sci. **44**, 1261–1267.

MacLennan, D. N. (**1990**). "Acoustical measurement of fish abundance," J. Acoust. Soc. Am. **87**, 1–15.

Miksis-Olds, J. L., and Stokesbury, K. D. (**2006**). "An autonomous system for evaluating groundfish populations: The Ocean Groundfish Observatory," ICES Annual Science Conference, Maastricht, The Netherlands, ICES CM O:08 (Sept. 19–23, 2006).

Misund, O. A. (**1997**). "Underwater acoustics in marine fisheries and fisheries research," Reviews in Fish Biology and Fisheries **7**, 1–34.

Murphy, B. R., and Willis, D. W. (**1996**). *Fisheries Techniques*, 2nd ed. (American Fisheries Society Publications, Bethesda, MD).

Nunnallee, E. P. (**1983**). "Scaling of an echo integrator using echo counts, and a comparison of acoustic and weir count estimates of a juvenile sockeye salmon population," FAO Fisheries Report **300**, 261–268.

Ransom, B. H., Steig, T. W., and Nealson, P. A. (**1996**). "Comparison of hydroacoustic and net catch estimates of Pacific salmon smolt (*Oncorhynchus* spp.) passage at hydropower dams in the Columbia River Basin, USA," ICES J. Mar. Sci. **53**, 477–481.

Simmonds, E. J. (**2003**). "Weighting of acoustic-and trawl-survey indices for the assessment of North Sea herring," ICES J. Mar. Sci. **60**, 463–471.

Simmonds, J., and MacLennan, D. (**2005**). *Fisheries Acoustics Theory and Practice*, 2nd ed. (Blackwell, Oxford, UK) 437 pp.

Wu, Q., and Jeng, B. (**2002**). "Background subtraction based on logarithmic intensities," Pattern Recogn. Lett. **23**, 1529–1536.

Ultrasonically determined thickness of long cortical bones: Three-dimensional simulations of *in vitro* experiments

Petro Moilanen[a)] and Maryline Talmant

Université Pierre et Marie Curie - Paris 6, Laboratoire d'Imagerie Paramétrique, UMR7623, Paris, F-75005, France and CNRS, UMR7623, Paris, F-75005, France

Patrick H. F. Nicholson, Sulin Cheng, and Jussi Timonen

Department of Physics and Department of Health Sciences, University of Jyväskylä, Jyväskylä, FIN-40014, Finland

Pascal Laugier

Université Pierre et Marie Curie - Paris 6, Laboratoire d'Imagerie Paramétrique, UMR7623, Paris, F-75005, France and CNRS, UMR7623, Paris, F-75005, France

It was reported in a previous study that simulated guided wave axial transmission velocities on two-dimensional (2D) numerically reproduced geometry of long bones predicted moderately real *in vitro* ultrasound data on the same bone samples. It was also shown that fitting of ultrasound velocity with simple analytical model yielded a precise estimate (UTh) for true cortical bone thickness. This current study expands the 2D bone model into three dimensions (3D). To this end, wave velocities and UTh were determined from experiments and from time-domain finite-difference simulations of wave propagation, both performed on a collection of 10 human radii (29 measurement sites). A 3D numerical bone model was developed with tuneable fixed material properties and individualized geometry based on X-ray computed tomography reconstructions of real bones. Simulated UTh data were in good accordance (root-mean-square error was 0.40 mm; $r^2 = 0.79$, $p < 0.001$) with true cortical thickness, and hence the measured phase velocity can be well estimated by using a simple analytical inversion model also in 3D. Prediction of *in vitro* data was improved significantly (by 10% units) and the upgraded bone model thus explained most of the variability (up to 95% when sites were carefully matched) observed in *in vitro* ultrasound data. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2769619]

I. INTRODUCTION

The gap between guided wave theory and experimental measurements of guided waves in bones can be bridged by using numerical simulation of wave propagation as a tool. This requires development and validation of a numerical bone model that takes into account the real macroscopic structure and material properties of bone.

In a previous study (Moilanen *et al.*, 2007b) we introduced a two-dimensional (2D) numerical bone model based on fixed material parameters and individualized geometry, which was extracted from x-ray computed tomography (CT) reconstructions of human radius specimens. The finite-difference method was used for numerical simulation of wave propagation in this 2D bone geometry. It was shown that Lamb's A_0 (fundamental antisymmetric) plate model accurately fits the phase velocity of a guided wave measured in this 2D bone geometry. In addition, it was shown that, despite the strong simplification of the bone model, significant correlations were obtained for measured ultrasound parameters between simulated and physical *in vitro* experiments when the bone model was based on the real specimen geometry and when the source and receiver positions were carefully matched. It was suggested by this study that upgrading the bone model into three dimensions (3D) should improve prediction of real *in vitro* data. On the other hand, it has been reported that, due to the strong circumferential curvature of long bones such as the radius, it is preferable to replace the A_0 plate model by a tubular F_{11} (fundamental flexural) model for the inversion of 3D axial transmission guided wave data in order to maintain the accuracy of thickness prediction (Moilanen *et al.*, 2007a).

The aims of this study were to upgrade a 2D numerical bone model into 3D, so as to better describe realistic macroscopic geometry of long bones, and to use this new bone model (a) to investigate the effect of sample-specific variations in cortical thickness and outer radius of curvature on the accuracy and precision of using the F_{11} tube model (Gazis, 1959; Silk and Bainton, 1979) for fitting simulated *in vitro* ultrasound data, and (b) to evaluate the extent to which the 3D bone model explains the variability observed in real *in vitro* ultrasound data. To this end, the simulated ultrasound data for virtual bones were compared (a) to F_{11} model and true wall thickness and (b) to real *in vitro* ultrasound data.

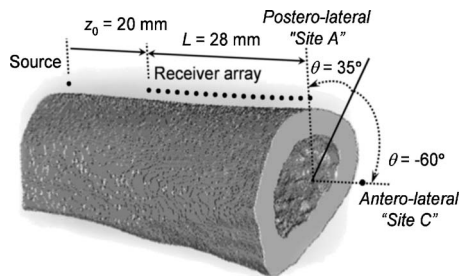a)Electronic mail: pemoilan@cc.jyu.fi

FIG. 1. Binarized 3D volume, as extracted from an x-ray CT reconstruction of human radius. Positions of source and receivers are shown.

II. MATERIALS AND METHODS

Simulations were performed using an extended 3D algorithm of the finite-difference code used in Moilanen *et al.* (2007b) and first introduced by Bossy *et al.* (2004).

Bone was considered as a homogeneous, transverse isotropic elastic medium. Its material parameters were considered fixed throughout the samples. Values of $c_L = 4000$ m/s and $c_T = 1800$ m/s for the axial direction, $c_L = 3410$ m/s and $c_T = 1645$ m/s for the transverse direction, and $\rho = 1.85$ g/cm^3 were used for bone, consistent with Bossy *et al.* (2004).

The actual macroscopic geometry was considered for each sample and was extracted from the 3D x-ray CT reconstructions used in Moilanen *et al.* (2007b). A binarized 3D volume and geometry of a simulation are illustrated in Fig. 1.

Temporal snapshots were used to visualize simulated wave field. Each simulation snapshot represents absolute values of the displacement velocity field, using a logarithmic gray scale combined with transparency effects to visualize bone. Illustration is given in Fig. 2. Four wave modes were observed in the simulated wave field, the first arriving signal (Wave 1), slower second wave (Wave 2), direct wave, and
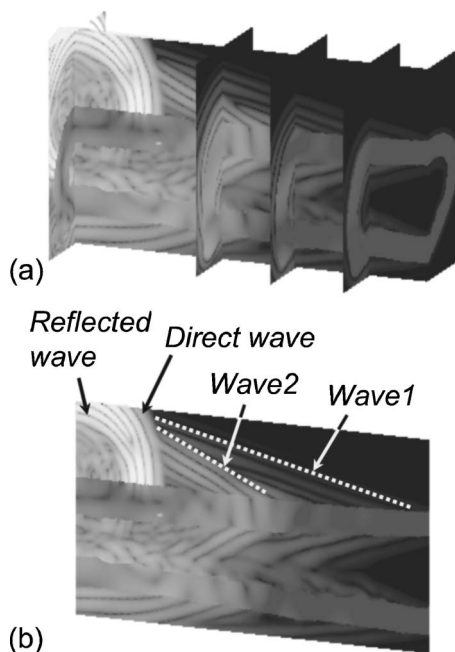

FIG. 2. Temporal snapshot of a simulation in 3D volume. (a) A number of longitudinal and transverse cross sections conjoined, and (b) bone longitudinal cross section alone. Propagating wave modes are illustrated.
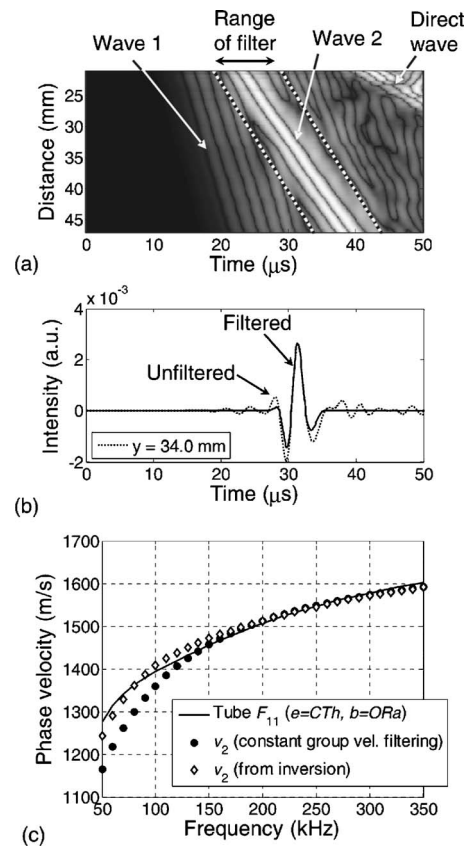

FIG. 3. Signal analysis: results for a simulated *in vitro* measurement. (a) Distance-time diagram, observed wave modes, and the range of group-velocity filter are shown. (b) Recorded (dashed line) and group-velocity filtered (solid line) signal. (c) Phase velocity of Wave 2 based on constant group-velocity filtering (solid markers), phase velocity of Wave 2 from the inversion scheme (open markers) and phase velocity of F_{11} mode for $e = $CTh and $b = $ORa (solid line).

reflected wave. Observation of snapshots indicated that wave fronts reached receivers at earlier time than return travel from the diametrally opposite site. Consequently, signals analyzed at the receivers came from transmission along the long axis of bone. Hence, it was justified to cut the 3D volumes in half in the horizontal plane and to use the upper half for actual simulation. This significantly reduced computing time and need for RAM memory of simulations.

Distance-time diagrams were used to illustrate signals measured at the array of receivers. This diagram illustrates absolute values of normal (perpendicular) particle velocity, using a gray scale, as functions of source-receiver distance and time history. Figure 3 illustrates such a distance-time diagram for a virtual bone sample. The recorded signal is composed of different contributions and one contribution, referred to as Wave 2, was used to assess bone thickness, according to the method introduced and reported in previous works [see, for example, Moilanen *et al.* (2006, 2007b)]. Wave 2 contribution was extracted from the whole signal, using the so-called method of group-velocity filtering, its phase velocity analyzed as a function of frequency and fitted by the dispersion curve of a given model. In Moilanen *et al.* (2007b), dedicated to 2D simulations, the model was the Lamb wave A_0; here, as 3D simulations were involved, the F_{11} dispersion curve, solution of the characteristic equation

Moilanen *et al.*: Ultrasonically determined cortical thickness

for a free cylindrical tube, was used. Output of the inversion scheme was both phase velocity of Wave 2 and estimation of cortical bone thickness UTh.

Reference cortical bone thickness and outer radius of curvature were deduced from the binarized simulation maps as average values within a 3D volume, with its longitudinal range limited by a range covered by the receiver array ($\Delta z = L = 28$ mm) and a circumferential range of $\Delta \theta = \pm 20°$ around the probe position (Fig. 1). The circumferential range chosen was approximately twice the diameter of a Fresnel zone covered by spherical wave at 1 mm distance from a point radiator (Pearce *et al.*, 2002). The thicknesses thus obtained represent average cortical thickness (CTh) and average outer radius of curvature (ORa) across the volume covered by the receiver array.

III. SIGNAL EXTRACTION AND THICKNESS ESTIMATION

Parameters used in group-velocity filtering (Appendix) were chosen after a study on reference waveguides, which were axisymmetric tubes with circular section, as the inversion scheme was based on F_{11} dispersion curve calculated for isotropic tubes.

Wave 2 extraction using group-velocity filtering requires the definition of certain signal processing parameters, including the value of group velocity used to move the window from one receiver position to a next one, capturing Wave 2. In Moilanen *et al.* (2007b), devoted to the 2D model and an inversion scheme based on A_0, a constant value of group velocity, time delay, and time width parameters allowed correct extraction of Wave 2. In 3D, setting these signal processing parameters to constant values was not as satisfactory in recovering F_{11} on isotropic tubes [Fig. 3(c)]. Therefore throughout this paper a varying group-velocity filtering was used.

On simulations of virtual isotropic tubes a discrepancy between the dispersion curve of Wave 2 and F_{11} was consistently observed, although F_{11} was calculated with material parameters, outer radius of curvature and thickness exactly identical to those used in simulation. Extraction of Wave 2 even failed for the lowest tube wall thickness tested ($e = 1$ mm) in the frequency range below 150 kHz. From the observation of known dispersion curves of guided modes in tubes, it was speculated that these difficulties in recovering F_{11} dispersion curve from the Wave 2 contribution in virtual tubes originated in the higher density of modes which propagate on tubes in comparison with the density of modes on plates in the same frequency range. In particular, interference between F_{11} and L_{01} (fundamental longitudinal mode) when their dispersion curves merge together was suspected to cause the difficulties in recovering F_{11}.

Using varying group velocity with regards to wall thickness and frequency based on group velocity of the F_{11} mode improved its extraction and thickness determination for virtual isotropic tubes as compared to using constant group-velocity parameter. Root-mean-squared error (RMSE) measured the agreement between phase velocity of Wave 2 and

TABLE I. Root-mean-square errors (RMSE, m/s) between phase velocities for different models across frequencies $f = 50 - 350$ kHz and all CTh ($n = 29$). For tube and F_{11} data: $e = $ CTh and $b = 7$ mm. Tube simulation data were spline-interpolated for $e = $ CTh from five (for isotropic) and nine (for anisotropic) simulations.

	$c_{ph}^{F_{11}}$ (isotropic tube)	Simulated v_2 (isotropic tube)	Simulated c_2 (anisotropic tube)
Simulated v_2 (isotr. tube)	37	⋯	45
Simulated v_2 (aniso. tube)	45	45	⋯
Simulated v_2 (aniso. bone)	54	63	70

F_{11} phase velocity. RMSE was 37 m/s for all frequencies and wall thicknesses in the range 2–5 mm (Table I).

While the bone model developed in this paper was considered anisotropic, the inversion scheme was always based on F_{11} calculated for an isotropic tube. Using simulations on virtual tubes with transverse isotropic elastic properties, we evaluated the error induced by this oversimplification. Material properties in the axial direction were identical in simulations and in F_{11} calculations (the compression and shear bulk wave velocities were, respectively, $c_L = 4000$ m/s and $c_T = 1800$ m/s). The outer radius was identical for virtual tubes and F_{11} calculations. RMSE, for all frequencies and thickness in the range 2–5 mm, was 45 m/s, i.e., 20% higher than the value of 37 m/s found for isotropic tubes (Table I). In addition, the bias in thickness determination caused by using the isotropic inversion model to fit data for anisotropic media was approximately $10 \pm 3\%$ (relative difference between UTh and true wall thickness). In this test case, the F_{11} dispersion curve was calculated for the exact value of outer radius used in the virtual tubes. However, it was observed that using a lower value of outer radius of curvature in F_{11} calculation ($b = 6.3$ mm) caused a 5% decrease in the accuracy error of UTh. Thus, some of the error in using an isotropic F_{11} model for inversion was compensated for.

In conclusion, Wave 2 was extracted within the inversion scheme using group velocity varying with frequency and wall thickness on the basis of F_{11} group velocity. The inversion model was based on F_{11} model calculated for isotropic tube with constant material properties and constant outer radius of curvature ($b = 6.3$ mm).

IV. RESULTS

A. Simulation results

For the center frequency ($f_c = 200$ kHz) and throughout all the virtual bone samples, phase velocity of Wave 2 (v_2) followed the same trend as phase velocity of F_{11} (Fig. 4). RMSE between Wave 2 phase velocity for virtual bones and F_{11} phase velocity (across all the samples and frequencies) was 54 m/s (i.e., less than 4% for v_2 of Table II).

Wave 2 phase velocity for virtual bones was also consistent with phase velocity of Wave 2 for simulations on

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

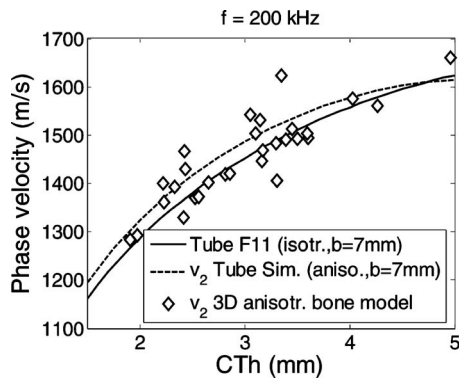Moilanen *et al.*: Ultrasonically determined cortical thickness   2441

FIG. 4. Simulated v_2 ($f=200$ kHz) for 3D bone geometry (diamond markers) and tubes ($b=7.0$ mm, dashed line, and closed square markers), and phase velocity for F_{11} tube model ($b=7.0$ mm, solid line).

isotropic and anisotropic tubes (data for an anisotropic tube is shown in Fig. 4), and RMSE were, respectively, 63 and 70 m/s (Table I).

RMSE between UTh and CTh data was 0.40 mm (for the whole data set). This corresponds, e.g., to 13% at CTh $=3$ mm. Linear regression yielded a highly significant correlation between UTh and CTh data ($r^2=0.79$, $p<0.001$) [Fig. 5(a)]. This result included exclusion of one outlier based on 95% confidence intervals of the first trial of linear regression, which originally yielded a correlation of $r^2=0.72$ ($p<0.001$). UTh underestimated CTh slightly, the relative difference between UTh and CTh being $-2\pm12\%$ [Fig. 5(b)].

When a UTh measurement with repeatability error of 5.1% (Table II) was used in a linear regression model as a predictor for true cortical thickness, the precision error of a single thickness estimate was $PE_{ThkEst}=0.42\pm0.02$ mm (i.e., $14\pm1\%$ for UTh$=3$ mm).

B. Simulated versus experimental data

Descriptive statistics of ultrasound variables determined for the simulated and experimental measurements are summarized in Table II. The rms coefficient of variation represents the uncertainty of positioning for simulated data and the repeatability error for experimental data.

Simulated ultrasound velocity, v_2, was significantly correlated with experimental v_2 [Fig. 7(a)], and variability in simulated data explained 90% of the variability observed in experimental v_2 data. The standard error of the estimate (SEE), i.e., root-mean-squared deviation of experimental v_2 data from the regression line, was SEE$=40$ m/s.

TABLE II. Descriptive statistics of measured ultrasound data ($n=29$).

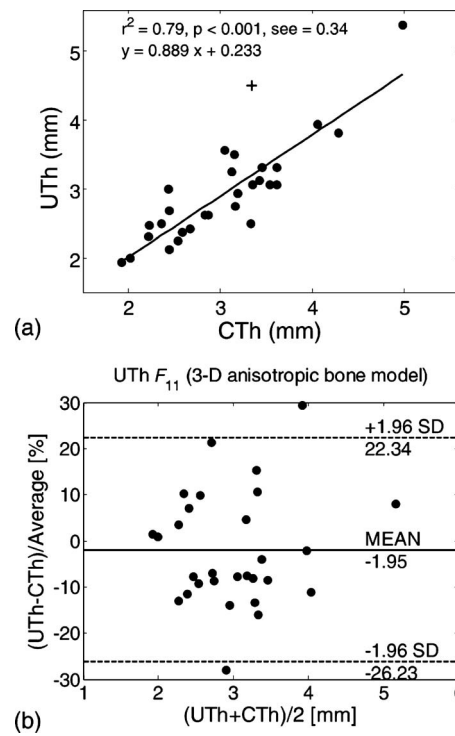|	Mean ±s.d.	Range	rms Cv (%)
Simulation results			
v_2 200 kHz (m/s)	1446±84	1285–1636	1.3[a]
UTh (F_{11}) (mm)	2.98±0.76	1.94–5.38	5.1[a]
Experimental results			
v_2 200 kHz (m/s)	1324±124	1099–1588	0.8
UTh (F_{11}) (mm)	2.19±0.71	1.23–4.16	3.5

Results for 2D simulations (Moilanen et al., 2007b).

FIG. 5. (a) Linear regression between UTh and CTh for 3D bone geometry. One sample (+) was excluded as an outlier based on the confidence intervals. Regression line fitted to the remaining ($n=28$) points. (b) Bland-Altman plot (Bland and Altman 1986) shows the relative accuracy error between UTh and CTh.

Accordingly, simulated and experimental UTh were significantly correlated, and simulated UTh data explained 91% of the variability in experimental UTh data [Fig. 7(b)] with SEE$=0.22$ mm.

V. DISCUSSION

A. Simulation results

According to preliminary studies on reference waveguides, using variable group-velocity filtering improved the consistency between v_2 and F_{11} phase velocity at frequencies $f<150$ kHz in particular. On the other hand, at $f=200$ kHz we observed that using variable group-velocity filtering implied only a slight improvement over using constant group-velocity filtering (results not given here). This suggested v_2 at $f=200$ kHz, obtained as an output of the inverse scheme, was not significantly affected by optimization but represented well the phase velocity of Wave 2. For UTh determination, on the contrary, using variable group-velocity filtering was necessary because the inversion was based on all the frequencies within the range investigated.

Conclusions obtained for reference waveguides cannot be extended to virtual bones, mainly due to interferences between ultrasonic modes, which differ for irregular geometries and rough interfaces. Study on reference waveguides just established the methodological limit of accuracy with an assumption that Wave 2 consistent with F_{11} can also be measured on bone geometry. Simulations on virtual bones were used to study the impact of realistic geometry.

The question of the real nature of Wave 2 was not, however, addressed in this paper or in Moilanen *et al.* (2007b). To answer this question we suggest to simulate propagation on a very large scanning length. This will give better resolution of the 2D-FFT transform and allow direct comparison with dispersion curves, without extraction (group-velocity filtering). We suggest also to characterize the profile of the particle displacement across the cortical wall to determine whether it is close to what is expected from a flexural wave. This is a work under progress. In this paper, we only addressed the question how good an agreement with F_{11} curve can be obtained, and we quantified this agreement.

The fact that, for F_{11}, the outer radius was fixed and a convenient value was used can appear artificial as the real outer radius of curvature of the bone samples varied, both within and between the samples. The agreement between v_2 and phase velocity of F_{11} has to be appreciated with respect to this parameter inserted in the inversion scheme. As, in the contrary, velocities for the two-dimensional A_0 Lamb model do not include the effect of an outer radius of curvature parameter, mismatch in adjusting this parameter can only yield uncertainty when fitting 3D tube models. Fixing the outer radius of curvature parameter could thus explain some of the lower accordance obtained here between v_2 (3D bone geometry) and phase velocity of F_{11} (RMSE=54 m/s) than obtained in Moilanen *et al.* (2007b) between v_2 (2D bone geometry) and phase velocity of A_0 (RMSE=33 m/s). In other words, this partly explains why the inverse scheme had a poorer performance in 3D than in 2D.

Final estimation for the precision on which a UTh measurement (with $PE_{ThkEst}=0.42$ mm devoted to the 3D inversion model and *in vitro* simulations represented in this paper) can discriminate changes in true cortical bone thickness is given by the least significant difference, which at the 95% confidence level is 2.8 $PE_{ThkEst}=1.20$ mm. The corresponding result reported for the 2D case was 0.70 mm (Moilanen *et al.*, 2007b).

It was expected that using the 3D inversion model (F_{11}) was nevertheless necessary for the accuracy of thickness estimation for 3D bone geometry, whereas the 2D inversion model (A_0) was suitable only for 2D bone geometry. This was demonstrated in Fig. 6 by using different inversion models for v_2 data in different bone geometries. Distributions of thickness data across the samples are shown.

B. Simulated versus experimental data

These results thus showed that development of a 3D bone model yielded a significant improvement, as compared to our previous 2D model (Moilanen *et al.*, 2007b), in explaining variability observed in real *in vitro* data. The agreement between simulated and experimental *in vitro* v_2 data increased from $r^2=0.79$ to $r^2=90$, and accordingly for UTh from $r^2=0.77$ up to $r^2=0.91$. The bone model developed thus includes the complete 3D geometry of bone cortex and any remaining discrepancy between simulated and experimental data was related to bone property variation and to the impact of bone marrow, which were the only remaining aspects not yet explained by the bone model. In addition, uncertainty of
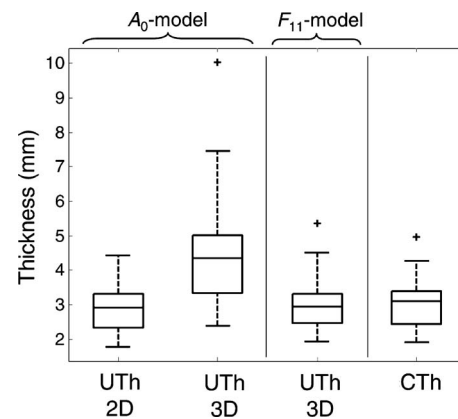

FIG. 6. Consistency between distributions of UTh data for 2D and 3D simulations in bone geometry when A_0 or F_{11} model was used for inversion. It is shown that using F_{11} inversion model is necessary for 3D bone geometry to maintain accuracy of cortical thickness prediction. CTh is shown for reference. Boxplots show the median, standard deviation, and 95% confidence interval of the median.

probe positioning between simulations and experiments and repeatability error of experiments (rms CV, Table II) contributed to discrepancy between the bone model and experimental data. Definition of thickness (based on the threshold level used for binarization of simulation volumes) caused some accuracy error between simulated and experimental data (see Table II), but this was not expected to affect significantly the precision of simulations.

As the bone model already explained up to 90% of the variability in experimental *in vitro* data, the role of positioning uncertainty and repeatability error becomes significant within the remaining 10% of unexplained variability. To this end, the impact of positioning uncertainty and repeatability
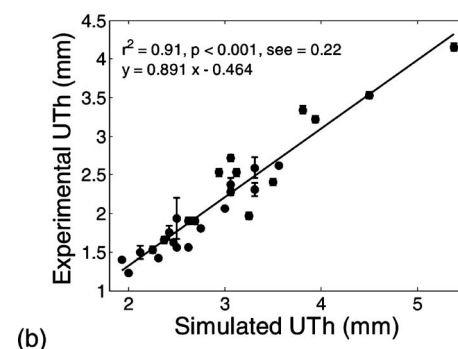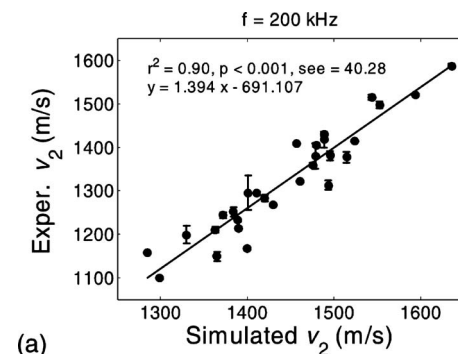

FIG. 7. Linear regression between experimental and simulated (a) v_2 data and (b) UTh data ($n=29$).

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Moilanen *et al.*: Ultrasonically determined cortical thickness   2443

TABLE III. Correlations between initially perfectly matched experimental and simulation results with randomized errors. rms Cv considered as the precision error of simulations and experiments.

	R^2	s.d. (R^2)	p
v_2 200 kHz (m/s)	0.95	0.02	$<1E-4$
UTh (F_{11}) (mm)	0.95	0.02	$<1E-4$

error on r^2 values between simulated and experimental data was evaluated numerically on initially perfectly correlated sets of variables ($r^2=1$, RMSE$=0$). The uncertainties of each point were individually randomized based on the normal distribution with standard deviation corresponding to uncertainties quantified for ultrasonic results (rms CV, Table II). The results were obtained after a thousand iterations (Table III). The impact of uncertainties was obtained as a reduction of these correlations, expressed by $1-r^2$. Thus, in general, uncertainties explained 4%–7% of the variability between simulated and experimental data. Given the error bounds, we thereby explained $95\pm2\%$ (for both v_2 and UTh) of the variability between simulated and experimental data. The remaining 5% could thus be explained by variation in bone material properties and by possible variation in impact of bone marrow.

In this section we did not discuss the accuracy of the bone model because it was expected that the choice of threshold level in the bone model's binarization and/or the choice of elastic parameters would introduce bias in the bone model, and this bias could only be compensated by iteration of simulations (given that, e.g., the true elastic parameters are known accurately enough). Such an iterative process requires a lot of computing time despite the relative efficiency of the finite-difference code used. Typical 2D simulations (Moilanen *et al.*, 2007b) required only a few dozens of megabytes of memory and less than 1 min of computing time on a 3 GHz desktop PC, whereas the corresponding 3D simulations of this paper were computed on an IBM RS/6000 computer and required approximately 5 to 6 h of computing time and 800–1200 Mbytes of RAM memory. Thus iteration of simulations was not possible within the framework of this paper. In this work we were focused on modeling realistic variability between simulations and experiments and the choices made in fixed reference parameters were not expected to affect significantly the precision of the bone model.

VI. CONCLUSION

We developed a 3D numerical bone model and used it to investigate the impact of realistic bone geometry on ultrasonic cortical thickness evaluation. It was shown that, despite geometrical variation within and between samples, and between samples and axisymmetric tubes, the low-frequency axial ultrasound velocity v_2 can be modeled by simple F_{11} tube theory with acceptable error bounds. Thus, ultrasonically determined thickness (UTh) based on fitting with the F_{11} model provides a good estimate of cortical bone thickness in 3D. In particular, we showed that accuracy of UTh as a predictor of true cortical thickness was mostly limited by parametrical model simplification as the inversion model

used did not account for variability in outer radius of curvature of bone shaft. For this reason we did not explain true thickness as precisely in 3D as we did in 2D by using Lamb's A_0 plate model, which did not include any parametrical simplification.

The numerical bone model also gave predictions of Wave 2 velocity in excellent agreement with experimental data, although experimental data were influenced by bone material property variations not taken into account by the 3D bone model. The advantage of developing a numerical bone model was that it enabled analysis of sensitivity independently for different aspects of bone property variation. In this work we showed in particular that modeling natural variability in bone geometry played the major role in modeling physical experiments. We also introduced an upgrade to our previous 2D bone model and the results reflect significant improvement in model prediction. Together this all demonstrates the power of numerical simulation as a tool for gaining detailed insight into understanding the sensitivity of ultrasonic parameters to different structural and material properties.

Future work should include investigation of the sensitivity of axial ultrasound parameters to bone material properties, such as stiffness, microporosity, and anisotropy. This could yield a multiparametric ultrasonic assessment, wherein different bone properties could be linked to a set of ultrasonic velocities for different wave modes by a solution of inverse problem such as used here to deduce thickness from one ultrasonic velocity. On the other hand, the clinical *in vivo* application should be adequately modeled. In particular, overlying layers of soft tissues are known to have significant impact on waveguide problems and therefore need to be modeled and their role understood. For both these aims, numerical simulation of transient wave propagation provides a powerful tool, and the bone model introduced by the present study provides a solid framework to investigate the impact of incorporating different, realistic, bone properties.

ACKNOWLEDGMENTS

APPENDIX: GROUP VELOCITY FILTERING

The purpose of group-velocity filtering method is to extract a part of a signal in time domain consistent with the energy propagation velocity of a given mode, and thus to enhance the ability of two-dimensional fast Fourier transform to separate the mode and determine its phase velocity (Moilanen *et al.*, 2006).

The method of (constant) group-velocity filtering introduced in Moilanen *et al.* (2007b) was based on fixed and constant group-velocity parameter, c_{gr}, as this was found suitable for extracting Wave 2 consistent with the A_0 plate mode and enabled computational simplicity. However, the group velocity of the F_{11} tube mode varies significantly with regards of frequency and wall thickness, within the range of frequencies analyzed. Wave 2, when consistent with F_{11}, can

only be extracted correctly by using variable $c_{gr}=c_{gr}(f_i,e_j)$ parameter, where f_i is the frequency, e_j wall thickness and $i=1,\dots,N$ and $j=1,\dots,M$ are indices. The varying group-velocity parameter is deduced from the theoretical group velocity of F_{11} mode for given parameters $f=f_i$ and $e=e_j$. Because in this study the wall thickness was considered unknown, variable group-velocity filtering was performed as part of the inverse scheme (Moilanen *et al.*, 2007b). As a result, using varying group-velocity filtering was influenced by optimization, whereas the constant filtering method yielded phase-velocity data independent of inversion scheme. The constant and varying group-velocity filtering methods did not thus lead generally to the same determination of phase velocity. In practice, however, using variable group-velocity filtering enhanced capturing Wave 2 and the two methods yielded relatively consistent determination of phase velocity [Fig. 3(c)].

Group-velocity filtering also includes parameter time width t_w and time delay t_d (Moilanen *et al.*, 2006), which were chosen fixed and constants throughout this paper. It is, however, technically possible to choose these parameters variable and adjust them by optimization. This may be necessary, for instance, to account for any signal delays additional to the theory based group-velocity delay. In a clinical application, such a delay could be propagation through overlying soft tissues of unknown thickness.

Bland, J. M., and Altman, D. G. (**1986**). "Statistical methods for assessing agreement between two methods of clinical measurement," Lancet **1**, 307–310.

Bossy, E., Talmant, M., and Laugier, P. (**2004**). "Three-dimensional simulations of ultrasonic axial transmission velocity measurement on cortical bone models," J. Acoust. Soc. Am. **115**, 2314–2324.

Gazis, D. C. (**1959**). "Three dimensional investigation of the propagation of waves in hollow circular cylinders, I. Analytical foundation," J. Acoust. Soc. Am. **31**, 568–573.

Moilanen, P., Nicholson, P. H. F., Kilappa, V., Cheng, S., and Timonen, J. (**2006**). "Measuring guided waves in long bones: Modelling and experiments in free and immersed plates," Ultrasound Med. Biol. **32**, 709–719.

Moilanen, P., Nicholson, P. H. F., Kilappa, V., Cheng, S., and Timonen, J. (**2007a**). "Assessment of the cortical bone thickness using ultrasonic guided waves: Modelling and in vitro study," Ultrasound Med. Biol. **33**, 254–262.

Moilanen, P., Talmant, M., Nicholson, P. H. F., Cheng, S., and Laugier, P. (**2007b**). "Ultrasonically determined thickness of long cortical bones: Two-dimensional simulations of in vitro experiments," J. Acoust. Soc. Am., in press.

Pearce, J., and Mittleman, D. (**2002**). "Defining the Fresnel zone for broadband radiation," Phys. Rev. E **66**, 056602, 1–4.

Silk, M. G., and Bainton, K. F. (**1979**). "The propagation in metal tubing of ultrasonic wave modes equivalent to Lamb waves," Ultrasonics **17**, 11–19.

Changes in signal parameters over time for an echolocating Atlantic bottlenose dolphin performing the same target discrimination task

Stuart D. Ibsen,[a] Whitlow W. L. Au, Paul E. Nachtigall,
Caroline M. DeLong, and Marlee Breese
*Marine Mammal Research Program, Hawaii Institute of Marine Biology, P.O. Box 1106 Kailua,
Hawaii 96734-1106*

This study documents the changes in peak frequency, source level, and spectrum shape of echolocation clicks made by the same dolphin performing the same discrimination task in 1998 and in 2003/2004 with spherical solid stainless steel and brass targets. The total average peak frequency used in 1998 was 138 kHz but in 2003/2004 it had shifted down nearly 3.5 octaves to 40 kHz. The total average source level also shifted down from 206 dB in 1998 to 187 kHz in 2003/2004. The standard deviation of these parameter values within time periods was small indicating a consistent difference between time periods. The average parameter values for clicks used when exposed to brass versus steel targets were very similar indicating that target type did not greatly influence the dolphin's average echolocation behavior. The spectrum shapes of the average clicks used in 1998 and in 2003/2004 were nearly mirror images of each other with the peak energy in 2003/2004 being concentrated where the 1998 clicks had the lowest energy content and vice versa. Despite the dramatic differences in click frequency content the dolphin was able to perform the same discrimination task at nearly the same level of success. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2772213]

I. INTRODUCTION

Much work has been done in the past to characterize the different clicks that are used by dolphins while performing echolocation tasks. There can be a large variation in the various parameters of the clicks, such as source level and peak frequency, between species, between individuals within the same species, and even between clicks of the same individual (Au, 1993). However, by averaging clicks together an idea of the general click behavior begins to emerge. It is this average click that has been used to understand the echolocation behavior of dolphins and allows for comparisons between individuals. The average click can also be used to study how an individual changes its echolocation behavior in different environments.

By looking at average click behavior it has been shown that cetaceans can make gross adjustments to the frequency content of their sonar signal when exposed to different levels of background noise. Such is the case with beluga whales which have been shown to shift the peak frequency of their clicks upwards due to background noise level changes (Au *et al.*, 1985). An individual whale produced clicks with peak frequencies between 40 and 60 kHz in San Diego Bay, but then produced clicks with peak frequencies between 100 and 120 kHz in Kaneohe Bay where the background noise was 12–17 dB greater. This upward shift in peak frequency also occurred along with an increase of up to 18 dB in click intensity.

It has been demonstrated that dolphins can be trained to make gross adjustments to their average click peak frequency while performing simple target detection tasks. A *Tursiops truncatus* was successfully trained to emit low frequency clicks with peak frequencies below 60 kHz when given a specific audio cue and to emit high frequency clicks above 105 kHz when given a different audio cue (Moore and Pawloski, 1990). The dolphin had the same level of success in determining the presence or absence of the target with the high frequency clicks as with the low frequency clicks. These observed changes in frequency content were not a natural behavior of the dolphin. The echolocation task was also a simple target present or absent determination which did not require any fine frequency content discrimination.

One aspect that has not yet been looked at is the natural stability of the average click used by an individual dolphin over a time period of years while performing the same fine target discrimination task in an environment with a nearly constant background noise level. Observing how the same dolphin might voluntarily change its average click over time in such a situation can provide insight into what cues might be used for these discrimination tasks. This study focuses on documenting the changes in the average clicks produced by a female bottlenose dolphin (*Tursiops truncatus*) during two similar target discrimination experiments conducted in 1998 and 2003/2004. These discrimination tasks were not simple target present or absent tasks, but rather more difficult dis-

---
[a]Current affiliation: University of California San Diego, Serf Bldg. Room 295 0435,9500 Gilman Drive, La Jolla, CA 92093. Electronic mail: sibsen@ucsd.edu
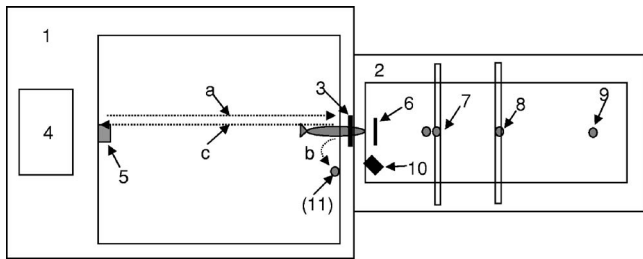
FIG. 1. The experimental setup for the phantom system. The different components are referred to in the text.

criminations between two targets with similar sound reflection characteristics. This information combined with information from additional experiments will set the stage to understand the implications of these findings as they pertain to possible discrimination cues and dolphin echolocation strategies.

II. MATERIALS AND METHODS

A. Subject and experimental conditions

The clicks used in this analysis were collected during experiments conducted at the Hawaii Institute for Marine Biology in Kaneohe Bay, Oahu, HI. The subject was an adult female Atlantic bottlenose dolphin (*Tursiops truncatus*), named BJ, who in 1998 was 13 years old. She was housed in a floating wire-net enclosure in Kaneohe Bay.

The experimental setup is shown in Fig. 1 and was the same for the 1998 and 2003/2004 experiments. The experimental enclosure consisted of two parts. The first was a floating pen frame, 8×10 m, that had a wire-net bottom (1) and was used to house BJ. The frame consisted of 55 gallon barrel floats underneath a wooden framework. The second was a target pen (2) that did not have a wire bottom and was used to support the recording hydrophone as well as the targets. The wire-net bottom was omitted from the target pen to prevent the production of extra confounding echoes during echolocation. BJ would station in a hoop (3) so that her head location was known and so she would have unobstructed echolocation access to the target pen (2). The computer system was set up in an electronics shack (4). The experimenters remained in the shack for the entire session. Each click that BJ produced was recorded by a hydrophone (7) that was located 2.5 m from the hoop. Clicks were recorded for all trials. An underwater camera (10) was set up to monitor BJ's stationing in the hoop to assure uniformity of her body orientation with respect to the hoop between trials. The hoop (3), recording hydrophone (7), transducer (8), and targets (9) were all at a depth of 1 m and all four parts were arranged to ensure all clicks recorded from BJ were on axis.

B. Experimental procedure

The procedure was the same for both the 1998 and 2003/2004 experiments. Each trial was started with a hand signal to cue BJ to swim from the stationing pad (5) to the hoop (3) along the dotted path (a) into the stationing hoop. The target would be lowered into the water with minimal splash, and BJ

would begin echolocating when the movable baffle (6) was lowered out of the way to provide an acoustically unobstructed path to the target pen (2).

A go/no-go paradigm was used in which the go response was associated with the standard steel target and the no-go response was associated with the comparison brass target. For the go response BJ backed out of the hoop and followed path (b) to touch a paddle and then returned to the shack along path (c). For the no-go response BJ stayed in the hoop for the duration of the 6 s trial. If BJ made a correct response, the trainer bridged her with a whistle and she was given a fish reward after returning to the stationing pad (5). If BJ made an incorrect response, she was recalled to the stationing pad (5) without a fish reward.

C. Click recording hardware and software

The click recording equipment for the 2003/2004 experiments used two hydrophones aligned such that one was directly behind the other when facing the dolphin. The two hydrophones were 5 cm apart and the time delay between the signal arriving at the first hydrophone and the second was enough to trigger the system to record the entire click with the second hydrophone. The first hydrophone was used only as a trigger for the system and did not record any information. The signal from the triggering hydrophone was amplified 56 dB and sent to a Measurement Computing Corporation PCI-DAS4020/12 analog to digital data aquisition board, set to a dynamic range of ±5 V, which digitized the signal at a sample rate of 1 MHz and triggered the system to begin recording with the second hydrophone. The click was then recorded in its entirety by the second hydrophone, amplified 36 dB, and then sent to the second channel of the same board using the same sampling rate of 1 MHz. No low pass filters were used because the high sampling rate itself prevented aliasing. A custom written LABVIEW program was used to control the data acquisition board and record each click straight to the hard drive of the computer for future analysis. This click recording equipment was based on the device used to collect clicks during the 1998 experiments, which is discussed in Aubauer and Au (1998).

D. Targets

There were two main targets: a solid brass sphere and a solid stainless steel sphere (both 7.62 cm in diameter). The steel sphere was considered the standard target and the brass was the comparison. These two targets comprised 92% of the target presentations. The other 8% of the target presentations were phantom targets. The generation of the phantom targets is covered thoroughly in Aubauer and Au (1998), but is briefly described here. Essentially the phantom targets were computer-generated targets whose acoustic reflection properties simulated those of the two real targets described earlier. The computer system calculated the echo from either the brass or steel spheres based upon the dolphin click that was recorded and played it back to the dolphin through a transducer. The system was fast enough to generate an echo specific for each incoming click. These phantom targets were presented to the dolphin in both the 1998 and 2003/2004

Ibsen *et al.*: Changes in signal parameters over time   2447

time periods as four nonreinforced phantom probe trials per session, two phantom steel and two phantom brass (Aubauer *et al.*, 2000). The probe trials were designed to understand if BJ would classify the phantom targets, especially the phantom steel, as the actual real standard target or as comparisons. The clicks made during these probe trials were included in the analysis because the target reflection characteristics of the real and phantom targets were very similar and because BJ made the same response to the phantom targets that she made to the real targets (e.g., she made the go response for phantom steel and real steel targets and the no-go response for the phantom brass and real brass target) (Aubauer *et al.*, 2000).

Each session consisted of 50 trials. Half of the trials were the standard steel target presentations (23 real steel and 2 phantom steel) and the other half of the trails were comparison brass target presentations (23 real brass and 2 phantom brass). The order of stimulus presentation was determined using a standard Gellermann series (Gellermann, 1993).

E. Click analysis

Custom designed and written programs were created to analyze the clicks from both the 1998 and 2003/2004 experiments separately. The analysis programs calculated the amplitude spectra for each click and determined the peak frequency. The peak-to-peak source level for each click was also determined based upon the known distance of 2.5 m between the dolphin and the recording hydrophone. The peak-to-peak source levels were reported in units of dB re 1 μPa at a reference distance of 1 m.

A large amount of variation was found in the peak frequency and source level between individual clicks making direct comparison between groups of clicks difficult. To develop a concept of the general trend that the clicks were taking it was necessary to average the clicks together in groups allowing direct comparison of the general behavior of the clicks made during exposure to different targets and at different times. In order to determine the general click behavior, clicks from each session were separated into those made during exposure to the (real and phantom) steel target and those made during exposure to the (real and phantom) brass target. The peak frequency values and source levels of each click made during exposure to the steel target were averaged together to yield the average peak frequency and average source level for the steel for each individual session. The same was done for the clicks made during exposure to the brass targets for each individual session. This allowed for direct comparisons of clicks between sessions within the same time period.

The averaging process was taken one step further by averaging the peak frequency values and source level values from all clicks made in 1998 and again separately in 2003/2004. These total average values allowed for a general comparison between both time periods. Averaging was also performed on all the clicks made during exposure to steel targets in 1998 and separately in 2003/2004 yielding a steel average for each time period. The same was done for the clicks made during exposure to the brass targets yielding a brass average.

A total of 4 sessions (200 trials) were analyzed from 1998 and 4 sessions (200 trials) from 2003/2004. Even though each time period contained the same number of sessions and trials, BJ used twice as many clicks in 1998 as she did in 2003/2004. Since the focus of this study is to compare BJ's click behavior between the two time periods the number of clicks were not equalized between the two time periods because it was not possible to choose which clicks to include and which to exclude from the analysis without possibly skewing the results. To maintain the focus on general click behavior it was more important to maintain a constant number of trials between both time periods than to maintain a constant number of clicks. Therefore, the total average values reported for each time period contain different numbers of clicks but reflect the general click behavior over entire sessions.

III. RESULTS AND DISCUSSION

The results from the above-described analysis are summarized in Tables I and II. As can be seen, there was a high degree of consistency with the average click peak frequency and source level between sessions within their respective time periods. In 1998 the total average peak frequency was 138 kHz with a standard deviation of 7 kHz, just 5% of the average. In 2003/2004 the total average peak frequency was 40 kHz with a standard deviation of 2 kHz, also just 5% of the average. However, between the 1998 and 2003/2004 time periods the total average peak frequency dropped 98 kHz, nearly 3.5 octaves.

Not only did the peak frequencies change but so did the entire spectral shape. The average click amplitude spectra are shown in Fig. 2 for six normalized characteristic clicks from 1998 and six normalized clicks from 2003–2004. The clicks were first normalized before averaging so the spectral shapes could be compared directly without intensity differences causing additional distortions because the interest is to analyze just the spectral shapes. As can be seen, the 2003/2004 clicks peak in the region where the 1998 clicks have the least amount of energy. They are nearly mirror images of each other in terms of where BJ concentrated the energy. The error bars show the upper and lower limits of the variability between the individual clicks. Even with the large amount of variability from click to click the inverse nature of the shapes is still evident. The average 1998 amplitude spectra does not peak at a relative normalized amplitude of one because the individual peaks differed by a few kilohertz from one another and ended up decreasing the peak value of the average frequency content. The 2003/2004 clicks all consistently peaked at 40 kHz causing the average spectra to peak at a value of one. The time domain wave form representation of characteristic echolocation clicks from both time periods is shown in Fig. 3 and reflect the frequency content differences described earlier.

Despite the dramatic differences in click frequency content, BJ was able to perform the same discrimination task at 92% correct in 1998 and at 98% correct in 2003/2004, nearly

2448   J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Ibsen *et al.*: Changes in signal parameters over time

TABLE I. A summary of the peak frequencies and source levels of clicks collected in 1998.

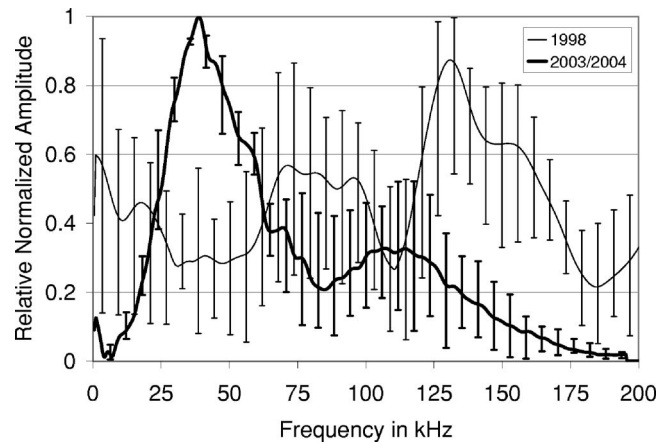| Date of data collection	Target type	Average peak frequency (kHz)	Average source level (dB)	Number of clicks analyzed
15 July 1998	Brass	145	205	996
	Steel	133	205	1202
15 July 1998	Brass	143	205	1187
	Steel	138	206	1334
24 July 1998	Brass	142	207	1675
	Steel	123	209	1431
13 August 1998	Brass	139	209	1576
	Steel	144	209	1285
Standard deviation:		7	2	Total: 10 686
Total average:		138	206	
Steel average:		135	207	
Brass average:		142	207	

FIG. 2. Average echolocation click amplitude spectra from 1998 and 2003/2004. Six characteristic spectra from each time period have been normalized and averaged together to produce these spectra. The error bars show the upper and lower limits of the variation in spectral values for the respective time periods.

the same high level of success. Unless BJ used time separation pitch to perform the discrimination, it is probable that she used different reflection cues from the targets in both time periods because the clicks were so different. The ability to use such different click types indicates BJ might have been able to change the acoustic cues she used to discriminate the targets. The ability to adaptively change the cues used to discriminate targets would be of high adaptation value from a survival standpoint when faced with changing and sometimes unpredictable acoustic surroundings. Future publications will present additional research and results that when combined with these results will shed more light on the possible acoustic and echolocation strategies BJ might have been using to perform these discrimination tasks.

It is interesting to note that there were no large differences in the peak frequencies or source levels between the clicks made when exposed to steel or to brass. The average steel peak frequency in 1998 was 135 kHz while for brass it was 142 kHz, which is only a 5% difference. The average steel peak frequency in 2003/2004 was 40 kHz, while for

brass is was 39 kHz. This indicates that the target type most likely did not influence BJ's echolocation behavior.

The source levels of the average clicks made by BJ were also very consistent within time periods but had large variation between time periods. The total average source level in 1998 was 206 dB re 1 μPa at a reference distance of 1 m with a standard deviation of only 2 dB. The total average source level in 2003/2004 was 187 dB with a standard deviation of only 2 dB, a drop of 19 dB. This indicates that the higher frequency clicks were produced by BJ mainly at the higher source levels. This increasing peak frequency with increasing source level is consistent with similar observations from other bottlenose dolphins and false killer whales (Au *et al.*, 1995; Houser *et al.*, 1999). This suggests physiological constraints that coarsely couple both source level and frequency content.

It is unknown what might have prompted BJ to change the average click she used to perform the discrimination task between both time periods. The same brass and steel targets were used in both experiments as well as the same experimental procedures and experimental setups. The high back-

TABLE II. A summary of the peak frequencies and source levels of clicks collected in 2003 and 2004.

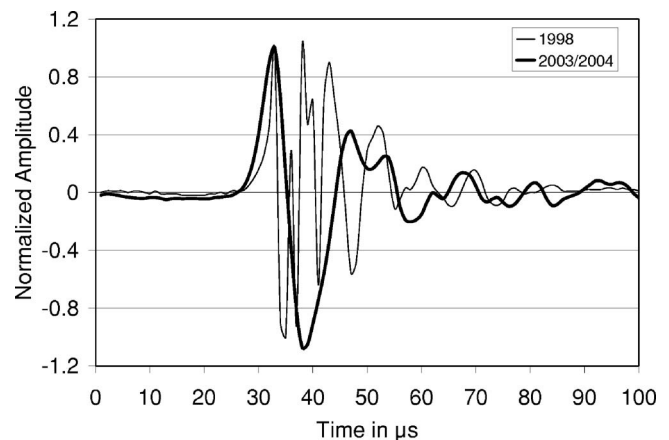| Date of data collection	Target type	Average peak frequency (kHz)	Average source level (dB)	Number of clicks analyzed
8 October 2003	Brass	41	187	498
	Steel	43	189	628
13 November 2003	Brass	36	183	439
	Steel	41	185	450
25 November 2003	Brass	38	189	627
	Steel	38	190	891
20 February 2004	Brass	40	188	667
	Steel	39	188	721
Standard deviation:		2	2	Total: 4921
Total average:		40	187	
Steel average:		40	188	
Brass average:		39	187	

FIG. 3. Characteristic echolocation click time domain wave forms from 1998 and 2003/2004. Both wave form intensities have been normalized for easier comparison.

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Ibsen *et al.*: Changes in signal parameters over time   2449

ground noise levels of Kaneohe Bay are largely caused by snapping shrimp whose activity did not have large changes between the two time periods as determined by occasional monitoring of the background noise levels in the bay. There is a tendency for an increase in snapping shrimp activity during the night, but the experiments were conducted during the day at about the same time.

One factor which might be involved is that BJ had a documented high frequency hearing loss when compared to other *Tursiops*. BJ's audiogram was measured using auditory evoked potential methods in 2001 and again in 2005 showing that she had an upper hearing limit of 45 kHz at both times with her most sensitive hearing being around 40 kHz. Audiograms from other *Tursiops* indicate that they can hear up to 150 kHz (Johnson, 1967; Au, 1993). With an upper hearing limit of 45 kHz, BJ would be able to hear less than 25% of the energy from the 138 kHz peak frequency clicks she made back in 1998. The most sensitive portion of her hearing range would not have been utilized with these clicks and in many cases there was a complete lack of energy in the click around this region. It seems there would have been little advantage for BJ to use the click amplitude spectra shapes from 1998 with an upper hearing limit of only 45 kHz. The clicks used in 2003/2004 were much better tailored for an upper hearing limit of 45 kHz because the average peak frequency in 2003/2004 was located right where her hearing was most sensitive at 40 kHz. She was able to perceive up to 42% of the energy in the 2003/2004 clicks making these clicks much more efficient.

Unfortunately, there are no audiograms for BJ before 2001 so her actual upper hearing limit in 1998 is unknown. However, the 1998 clicks appear to be much better suited for use with an upper hearing limit higher than 45 kHz. Perhaps the observed shift of the energy content of the clicks to lower frequencies between both time periods was an attempt by BJ to compensate for a possible high frequency hearing loss suffered between 1998 and 2003. Although the data presented here do not conclusively show that a hearing loss actually occurred and the possible cause of such a hearing loss in this specific case is not known, it has been shown that dolphins can lose their high frequency hearing. Possible factors that have been associated with reducing dolphin hearing ranges are age (Brill *et al.*, 2001; Ridgway and Carder, 1997) and long term residence in a noisy environment (Houser and Finneran, 2006).

Despite the large voluntary changes BJ made to her echolocation clicks, her ability to perform, the same task did not change. To better understand how BJ was able to do this, it is necessary to first understand the cues BJ used. The transfer function of the steel and brass targets show many subtle spectral differences in the way both targets reflect sound and any of these differences could have been potential cues. It is unknown which specific cues BJ might have been using to discriminate between the two targets and how those cues

might have been affected by her voluntary change of echolocation clicks. One way for future experiments to really explore this issue is to experimentally change the reflection characteristics of the targets used in the discrimination task and behaviorally determine if BJ can perceive those changes. This can be accomplished by systematically filtering the frequency content of the returned echoes using the phantom echo technology described in Aubauer and Au (1998). The results of these experiments can show what frequency bands BJ pays attention to while performing discrimination tasks and that information can potentially reduce the set of possible auditory cues.

Au, W. (**1993**). *The Sonar of Dolphins* (Springer, New York), pp. 32, 41–46, 66–70, 148–152.

Au, W., Carder, D., Penner, R., and Scronce, B. L. (**1985**). "Demonstration of adaptation in beluga whale echolocation signals," J. Acoust. Soc. Am. **77**, 726–730.

Au, W., Pawloski, J., Nachtigall, P., Blonz, M., and Gisner, R. (**1995**). "Echolocation signals and transmission beam pattern of a false killer whale (*Pseudorca crassidens*)," J. Acoust. Soc. Am. **98**, 51–59.

Aubauer, R., and Au, W. (**1998**). "Phantom echo generation: A new technique for investigating dolphin echolocation," J. Acoust. Soc. Am. **104**, 1165–1170.

Aubauer, R., Au, W., Nachtigall, P., E., Pawloski, D. A., and DeLong, C. M. (**2000**). "Classification of electronically generated phantom targets by an Atlantic bottlenose dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **107**, 2750–2754.

Brill, R., Moore, P., and Dankiewicz, L. (**2001**). "Assessment of dolphin (*Tursiops truncatus*) auditory sensitivity and hearing loss using jawphones," J. Acoust. Soc. Am. **109**, 1717–1722.

Gellermann, L. (**1993**). "Chance orders of alternating stimuli in visual discrimination experiments," J. Phys. Chem. **42**, 206–208.

Houser, D., and Finneran, J. (**2006**). "Variation in the hearing sensitivity of a population of determined with the use of evoked potential audiometry," J. Acoust. Soc. Am. **120**, 4090–4099.

Houser, D., Helweg, D., and Moore, P. (**1999**). "Classification of dolphin echolocation clicks by energy and frequency distributions," J. Acoust. Soc. Am. **106**, 1579–1585.

Johnson, S. (**1967**). "Sound detection thresholds in marine mammals," in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, New York), Vol. **2**, pp. 247–260.

Moore, P., and Pawloski, D. (**1990**). "Investigation of the control of echolocation pulses in the dolphin (Tursiops truncatus)," in *Sensory Abilities of Cetaceans: Laboratory and Field Evidence*, edited by J. A. Thomas and R. A. Kastelein (Plenum, New York), pp. 305–316.

Ridgway, S., and Carder, A. (**1997**). "Hearing deficits measured in some *Tursiops truncatus*, and discovery of a deaf/mute dolphin," J. Acoust. Soc. Am. **101**, 590–594.

Feature weighting in "chick-a-dee" call notes of *Poecile atricapillus*

Carly M. Nickerson, Laurie L. Bloomfield, and Michael R. W. Dawson
Department of Psychology, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

Isabelle Charrier
Bioacoustics Team, NAMC-CNRS UMR 8620, Université Paris Sud, Bat.446, F-91405, Orsay, France

Christopher B. Sturdy[a]
Department of Psychology, Centre for Neuroscience, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

Artificial neural networks were trained to discriminate between different note types from the black-capped chickadee (*Poecile atricapillus*) "chick-a-dee" call. Each individual note was represented as a vector of summary features taken from note spectrograms and networks were trained to respond to exemplar notes of one type and to fail to respond to exemplar notes of another type. Following initial network training, the network was presented novel notes in which individual acoustic features had been modified. The strength of the response of the network to each novel and shifted note was recorded. When network responses were plotted as a function of the degree of acoustic feature modification and training context, it became clear that modifications of some acoustic features had significant effects on network responses, while others did not. Moreover, the training context of the network also played a role in the responses of networks to manipulated test notes. The implications of using artificial neural networks to generate testable hypotheses for animal research and the role of context are discussed. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2770540]

I. INTRODUCTION

A perceptron is a simple artificial neural network (ANN) that consists of input processing units that encode stimuli, as well as output processing units that represent responses to these input patterns. These two sets of processing units are linked by a set of modifiable, weighted, connections. The weights are initially random, but with the use of a learning rule they achieve values that mediate a desired stimulus-response mapping (Rosenblatt, 1962; for a more general introduction, see Dawson, 2004).

Perceptrons are very simple ANNs because they do not have intermediate processing units (called hidden units) that intercept and modify input unit signals before they reach the output units. As a result, a perceptron's response is based solely upon the sum of the weighted signals from each input unit. In the absence of hidden units, there are many stimulus-response mappings that cannot be computed by a perceptron (Minsky and Papert, 1988). Because of this, many researchers prefer to use more powerful networks that include hidden units, such as the multilayered perceptron (e.g., Rumelhart *et al.*, 1986).

However, even though perceptrons cannot learn to compute every stimulus-response mapping, the subset of mappings that they can represent is informative and interesting. In particular, perceptrons have been shown to be important

models in the domain of discrimination learning (Dawson, 2005; Dawson and Spetch, 2005; Nickerson *et al.*, 2006; Yaremchuk *et al.*, 2005). The purpose of the current paper is to explore the utility of perceptrons in the context of a particular discrimination learning paradigm that explores the component notes of the "chick-a-dee" call of the black-capped chickadee (*Poecile atricapillus*).

A. Note discrimination by black-capped chickadees

Black-capped chickadees produce a chick-a-dee call that consists of four note types categorized as A, B, C, and D by bioacousticians (Ficken *et al.*, 1978; see Fig. 1 for exemplar note types). Each note type can be repeated or omitted, but the note types always appear in a fixed order within the call. For instance, A notes occur before B notes, B notes before C notes, and C notes before D notes. The fixed sequence in which these note types occur in the black-capped chickadee call is analogous to the syntax of human language (Hailman, 1985; Hailman and Ficken, 1986). That is, the call's generative syntax allows for countless note combinations, which parallels the human language system in the ability to generate an infinite variety of sentences from a finite vocabulary.

Pursuing this analogy, different combinations of notes in the chick-a-dee call might convey different messages, with each combination representing a different "sentence." In fact, Ficken *et al.* (1994) noted that the chickadee call syntax aids in species identity, mate choice, and territory defense, and the call composition differs according to context in Mexican

---

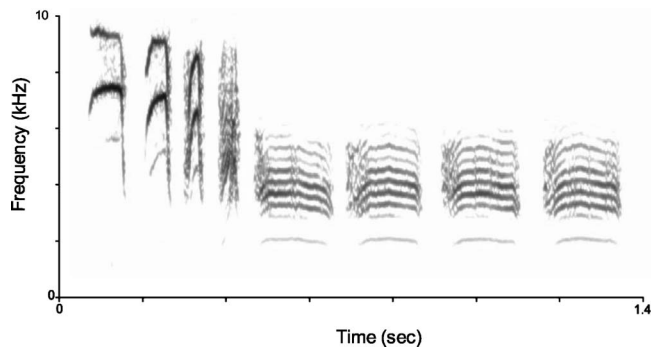[a] Electronic mail: csturdy@ualberta.ca

FIG. 1. Sound spectrogram of a "chick-a-dee" call (note types AABCD-DDD) with exemplars of A, B, C, and D note types.

chickadees (*Poecile sclateri*). As well, Charrier and Sturdy (2005) found that the black-capped chickadee fails to respond to atypical note ordering, and concluded that a fixed order is required to educe behavior. Finally, Templeton *et al.* (2005) showed that chick-a-dee calls and their notes vary with the perceived threat level of a predator. Clearly, accurate perception of chick-a-dee calls and note type composition are important to the survival of chickadees in the field.

Researchers have also studied the functions of the constituent notes in the chick-a-dee call, independently of note ordering within calls. Playback studies have indicated that the black-capped chickadee D note mediates information on species and flock identification (Mammen and Nowicki, 1981; Nowicki, 1983), while the C note reveals information about food location and availability (Freeberg and Lucas, 2002). In addition, individual notes within the call may offer information about the caller such as identity and gender (Charrier *et al.*, 2004). Statistical analyses have also shown that any constituent note (A, B, C, or D) carries enough information to determine whether the note was uttered by a black-capped chickadee or by another species, the mountain chickadee (*Poecile gambeli*; Dawson *et al.*, 2006).

Considering that both the syntax and the component notes of the chick-a-dee call appear to have functional roles, chickadees must be able to categorize these notes into note types (e.g., to process syntax). Indeed, black-capped chickadees can differentiate between the bioacoustically defined A-, B-, C-, and D-note types (Sturdy *et al.*, 2000). For instance, Sturdy *et al.* (2000) found that between-category learning was faster than within-category learning of these note types. This result has led to subsequent research that studies the specific features and mechanisms used by black-capped chickadees to perceive and subsequently categorize individual notes.

For instance, Charrier *et al.* (2005) more closely examined how black-capped chickadees categorized the non-D note types in an operant discrimination experiment using a GO/NO GO paradigm. Of particular interest was a second phase that was conducted after the birds had successfully completed the acquisition phase of the discrimination. Charrier *et al.* (2005) investigated note classification mechanisms by presenting trained birds with notes that had been acoustically manipulated. In particular, they took exemplar notes (A, B, and C) and shifted their spectral features up or down

by shifting the entire spectrogram of a note up or down a specific interval from its original starting frequency. Charrier *et al.* (2005) found that black-capped chickadees could accurately categorize note types based on frequency characteristics alone (e.g., start frequency and ascending frequency modulation), and that adjacent note types were misclassified when the notes were modified by these spectral features (e.g., A notes shifted down were confused with B notes, while B notes shifted up were confused with A notes). That is, shifting the entire spectrogram of a note produced a systematic pattern of misclassifications by the birds.

B. Note discrimination by perceptrons

While is was clear that overall note frequency controlled note-type classification, what remained to be determined was why chickadees miscategorize shifted notes in the particular fashion observed by Charrier *et al.* (2005). Nickerson *et al.* (2006) began to answer this question by creating a model of note discrimination. They trained perceptrons in a simulation of the Charrier *et al.* (2005) experimental paradigm. These perceptrons used nine input units to represent each note as a set of summary features taken from sound spectrograms. A single output unit was trained to turn on when the network was presented a note of one type (e.g., an A note) and to turn off when the network was presented a note of a different type (e.g., a B note). All of the discrimination learning conditions in the Charrier *et al.* (2005) study were successfully simulated in this fashion.

After training, featural representations of shifted notes were presented to the trained perceptrons, and the output unit responses were analyzed. Results from the perceptron response activities were compared to the classifications made by the black-capped chickadees, and strong parallels between the ANNs and the birds were found. For example, as A notes were shifted downwards in frequency both the chickadees and the perceptron responded as if they were B notes. Likewise, B notes shifted upwards in frequency were responded to as if they were A notes. Given the high degree of correspondence between perceptrons and chickadees when presented shifted notes, it appeared that the networks provided a solid first step in developing a model to determine how birds perform these classifications.

However, in both the Nickerson *et al.* (2006) simulation study, and in the Charrier *et al.* (2005) chickadee study, when novel notes were created by shifting frequencies, all of the frequency features were affected by the shift, and none of the duration features were modified at all. As a result, in these studies it was not possible to examine how individual summary features independently contributed to note-type recognition. This is unfortunate, because previous studies do indicate that different features are important for classifying different notes types. For example, Dawson *et al.* (2006) presented the non-D note types to a multilayer perceptron as a categorization task. They found that only four of the nine summary features were reliable identifiers of note type when considered across a number of predictor methods (ANN, multiple regression, and linear discriminant analysis).

There is another general and prevalent issue in animal learning and cognition that was not addressed by either the Nickerson *et al.* (2006) or the Charrier *et al.* (2005) study, namely the effect of learning context on performance. It has been well established that contextual cues have strong and pervasive effects on learning performance. For instance, as so elegantly predicted by Miller's (2006) comparative hypothesis, context, when salient, can control conditioned responding to such an extent that putative CSs are ignored by animals and the context of the conditioning apparatus itself controls responding. Further, Bouton's (2004) research on extinction has shown in a wide variety of experiments the effects that context can have on the recovery of a previously extinguished behavior.

Context has also been shown to be crucial to chickadee's processing of auditory stimuli. Lee *et al.* (2006) revealed that chickadees are sensitive to context when testing generalization in an absolute pitch discrimination task. Lee *et al.* (2006) found that birds tested with novel tones in the absence of training tones failed to maintain their previously learned discrimination and did not show evidence of transfer to the novel tones. In contrast, birds that were tested with both training and novel tones presented in the same session, showed solid evidence of generalization from training to novel tones. Clearly, as in other studies of learning, context plays a vital role in songbird auditory perception. The second aim, then, of the current research was to determine what, if any, were the effects of discrimination training context on perceptron performance both in the acquisition of the task, and in the generalization to novel, manipulated notes.

In summary, the purpose of the present simulation study is twofold. First, we wanted to more directly examine the contribution of each individual summary feature to note discrimination, including duration features. To do so we manipulated each acoustic feature independently in each note type in an effort to specify the features potentially used by birds to classify notes into types, and more specifically, to establish which features determine how the shifted notes were categorized by a perceptron. Second, we set out to determine whether learning context affected either acquisition or transfer to manipulated notes by the perceptrons. To accomplish our second aim, we included training contexts not used in either Charrier *et al.* (2005) or in Nickerson *et al.* (2006) and compared performance during initial training and during generalization tests.

II. METHODS

A. Training stimuli

Sixty notes (20 A, 20 B, and 20 C) previously utilized by Charrier *et al.* (2005) were used to create the training sets for the perceptrons. D notes were not included in this set because they are acoustically dissimilar from non-D notes, and cannot be represented using the same summary features used to represent non-D notes. In addition, D notes are rarely confused with other note types (Charrier *et al.*, 2004; Sturdy *et al.*, 2000).

The 60 training notes were randomly selected from 60 different calls recorded from birds from Colorado, and On-

tario, Canada. An equal representation of both locations was included, and all notes were of high quality. Call-note amplitude was equalized using SIGNAL v 4.0 (Engineering Design, Berkeley, CA). As well, each call was bandpassed filtered (1–10 kHz) to remove background noise.

Each note was represented as a vector composed of seven numerical values. These values were of seven summary features that have been used in several previous studies to represent chick-a-dee call notes (Dawson *et al.*, 2006; Nickerson *et al.*, 2006; Nowicki and Nelson, 1990). These features were start frequency (SF), peak frequency (PF), end frequency (EF), maximum (loudest) frequency (Fmax), ascending duration (AD), descending duration (DD), and total duration (TD). These particular features were selected because of their mutual independence. That is, any one of these features could be manipulated without affecting any of the other six features.

The spectral measurements (SF, PF, EF, Fmax) were quantified in digital spectrograms (window size 1024 points, frequency precision=43 Hz) using a cutoff amplitude of −35 dB relative to peak amplitude. We used the loudest harmonic of A and B notes (i.e., the carrier frequency), and the fundamental frequency of C notes in our measurements. The temporal measurements (AD, DD, TD) were likewise quantified in a digital spectrogram (window size 256 points, duration precision=5.8 Hz).

After obtaining the feature measurements, we converted each value into a z-score. This reduced the range of values presented to the input units of the perceptrons while maintaining the essential characteristics of the raw data (Dawson *et al.*, 2006).

B. Test stimuli

The 60 training stimuli were also used to create a large set of shifted stimuli (4200 notes) to be used to test a perceptron's responses after it had learned to discriminate training stimuli. Each training stimulus was used to create 70 different shifted notes. Each shifted note was identical to the training stimulus with the exception that the value of one of its features was shifted either upwards or downwards in value. Each feature was shifted by one of ten different amounts (±0.5, ±1.0, ±1.5, ±2.0, and ±2.5 s.d.). The 70 shifted variants of each note resulted by applying each of these ten shifts to each of the training note's seven features.

When shifted notes were created, they were created from the raw feature summary vectors. That is, they were not created from the z-score representation of note features. The approach to creating the shifted notes was analogous to that described by Charrier *et al.* (2005). The s.d. values for the shifts of each feature were dependent on note type, and were based on values previously reported by Charrier *et al.* (2004; see Table I). Once the shifted notes were created from the raw feature values, they were converted to z-scores using the means and s.d. of the original training stimuli.

C. Network architecture

All of the perceptrons trained in the simulations reported in the following consisted of seven input units (one for each

TABLE I. Provided here are the s.d. values that were used for shifting individual features according to note and feature type. The value was multiplied by each of the s.d. increments, and then added to or subtracted from the original value measured from the spectrogram of that note depending on direction of shift. For example, the SF s.d. values for an A note were (824 ×0.5), (824×1), (824×1.5), (824×2), and (824×2.5), and each of these products, added to and subtracted from the original measured A note SF value, gave the shifts across s.d.

Feature	A	B	C
SF (Hz)	824	518	144
PF (Hz)	381	660	938
EF (Hz)	631	451	126
Fmax (Hz)	397.6	647.3	1009
AD (ms)	13.1	5.7	5.1
DD (ms)	6.3	4.8	2.2
TD (ms)	17.4	8.5	4.4

note feature) and a single output unit. The output unit used the logistic equation to convert the sum of the weighted signals from the input units (the net input) into an activation value that ranged between 0 and 1. The logistic equation is a sigmoid-shaped function that is frequently used in ANNs to apply a nonlinear transformation to net inputs (e.g., Rumelhart *et al.*, 1986).

D. Network training on unshifted notes

The training set of 20 A, 20 B, and 20 C notes was used to replicate the training conditions used by Charrier *et al.* (2005). For example, the discrimination group A+B− was comprised of 20 A notes and 20 B notes. Perceptrons that served as subjects in this condition were trained to turn their output unit "on" when presented A notes and to turn their output unit "off" when presented B notes. This approach was used to create four different discrimination learning conditions that paralleled the four studied by Charrier *et al.* (2005): A+B−, A−B+, B+C−, and B−C+. Ten different networks served as "subjects" in each of these four conditions. Each subject was different from another because prior to learning, all of the connection weights in a perceptron were set to random values selected from the range between −0.1 and 0.1.

A gradient descent rule was used to train each perceptron (Dawson, 2004, 2005). This rule modifies connection weights in such a way that output unit error decreases as patterns are repeatedly presented. Connection weights were modified by this learning rule after each note was presented during training. Training proceeded by presenting patterns in "epochs." In a single epoch, the network was trained on each of the 40 stimuli being used for a particular condition. The order of pattern presentation was randomized every epoch. The learning rate for the simulations was set at 0.5. The training was conducted with the Rosenblatt program (Dawson, 2005).

Training proceeded until a perceptron generated a "hit" for each of the 40 stimuli that made up its training set. A hit was defined as the output unit producing activity of 0.9 or higher when the desired response to a pattern was on, or producing activity of 0.1 or lower when the desired response

to a pattern was off. Every perceptron that was trained in these simulations successfully learned the correct pattern of discrimination responses to all 40 patterns that it was presented.

E. Network testing on shifted notes

After a perceptron had successfully learned to discriminate one type of note from another (e.g., A+B−), it was then presented each of the 4200 shifted notes. Instead of training the perceptrons on these notes, we merely recorded the output unit's activity to each of these shifted stimuli. The responses of the perceptrons to these novel, shifted notes could then be used to explore the contribution of each individual feature to the overall responses that had previously been examined in birds (Charrier *et al.*, 2005) and in perceptrons (Nickerson *et al.*, 2006).

III. RESULTS

A. Training results

The networks trained in both AB discrimination learning conditions converged (i.e., successfully learned responses to all 40 input patterns) after an average of 818.5 epochs. The networks trained in both BC conditions converged after an average of 357.4 epochs. This difference is statistically significant ($t=365.31$, $df=18$, $p<0.001$), and is consistent with the fact that the visual appearance of the sound spectrograms (and thus the summary features) for A and B notes are more similar to one another than to the spectrograms of C notes. Because of this, it is not surprising to find that one discrimination task is harder to learn than another.

B. Testing results
1. Analysis of network responses to shifted notes

After discrimination training, each trained network was presented each of the 4200 shifted notes, and the activity produced in a network's output unit was recorded. These activities were then averaged over the ten perceptrons that served as subjects in each condition. These averaged responses were plotted as a function of shifted feature and note type, producing 21 different graphs. Each graph depicts how a network that was trained in a particular discrimination condition responded to one of the two note types presented in that set. These responses were plotted as a function of the shifts of one of the note's seven descriptive features. For instance, one graph would illustrate the responses of networks trained in the A+B− discrimination condition and other networks trained in the A−B+ condition when presented an A note that had its starting frequency (SF) shifted upwards and downwards.

One can easily convert the information derived from a graph to a brief qualitative description of the effects shifting the values of a note's feature. Table II provides a summary of these qualitative descriptions for all of the graphs that were produced in the analyses of network responses to notes with shifted features. The notation used in Table II indicates whether a feature shift resulted in a note becoming more like

TABLE II. A qualitative account of specific training conditions and the corresponding network generalizations to each note type with individually shifted features.

		SF −	SF +	PF −	PF +	EF −	EF +	Fmax −	Fmax +	AD −	AD +	DD −	DD +	TD −	TD +
A	A+B−	↓ A	···	↓ A	···	···	↓ A	···	↓ A	↓ A	···	···	···	↓ A	···
	A−B+	↑ B	···	↑ B	···	···	↑ B	···	↑ B	↑ B	···	···	···	↑ B	···
B	A+B−	···	↑ A	↓ A	↑ A	···	···	↑ A	···	···	↑ A	···	···	···	↑ A
	A−B+	···	↓ B	↑ B	↓ B	···	···	↓ B	···	···	↓ B	···	···	···	↓ B
	B+C−	···	···	↓ B	···	↓ B	···	↓ B	···	···	↓ B	···	↓ B	···	↓ B
	B−C+	···	···	↑ C	···	↑ C	···	↑ C	···	···	↑ C	···	↑ C	···	↑ C
C	C−B+	···	···	···	↑ B	···	···	···	↑ B	↑ B	···	···	···	↑ B	···
	C+B−	···	···	···	↓ C	···	···	···	↓ C	↓ C	···	···	···	···	↓ C

one of the note types (e.g., ↑ A indicates that a shift causes a note to be more A-like), less like one of the note types (e.g., ↓ A), or had no effect at all (---).

An examination of Table II reveals several different regularities. First, when a feature shift affected network responses, this effect depended upon the specific reinforcement condition that perceptrons were trained in. For example, when SF was shifted downward in an A note, this led to the note being less A-like for perceptrons trained in the A+B− condition, and to the note being more B-like for perceptrons trained in the A−B+ condition. A second property is that shifts of some features were more likely to affect network responses than were shifts of other features. For instance, shifts of PF produced changes in network performance in ten different cells of Table II, while modifications of DD only produced changes in two of the table's cells. A third property revealed by Table II is that the effect of a feature shift was context dependent, in the sense that the effect depended upon which two types of notes were involved in discrimination learning. For instance, manipulations of SF had effects on networks that were trained to discriminate A notes from B notes, but did not affect networks that were trained to discriminate C notes from B notes.

2. Analysis of network connection weights

Why do the perceptrons respond in the fashion that is summarized in Table II? Ultimately, the responses of a perceptron are governed by the weights of the connections between the input units and the output unit. An input feature that has a strong excitatory influence on the output unit will have a large positive weight. An input feature that has a strong inhibitory influence on the output unit will have a large negative weight. An input feature that has little influence on the output unit will have a near-zero weight. An examination of the connection weights of the trained perceptrons should provide insight into the regularities that are apparent in Table II.

Table III presents the connection weights for each discrimination learning condition averaged over the ten perceptrons that served as "subjects" in the condition. The pattern of weights presented in Table III is consistent with the pattern of network responses that was summarized in Table II, as is detailed in the following.

First, Table III shows that SF and Fmax were important when differentiating between A and B note types because they had the largest absolute values of connection weights in the AB discrimination learning conditions. Similarly, AD and Fmax were the most important features used to differentiate between B and C note types. In general, though, Fmax and PF were the most significant indicators of note type because they had large connection weights in all of the discrimination learning conditions.

Second, several summary features were associated with near zero connection weights, and therefore had little importance for differentiating between note types. This was particularly true for DD in the AB note discrimination conditions and for SF in the BC note discrimination conditions. Overall, DD was the least relevant feature in differentiating between note types.

Third, there was a distinct reflection of connection weights depending upon which note was reinforced in a particular discrimination learning condition. For example, con-

TABLE III. Connection weight average values and their standard deviations in each training set for the association between an input unit and the output unit.

	SF	PF	EF	AD	DD	TD	Fmax
A+B−	7.86 (0.032)	3.47 (0.026)	−1.73 (0.023)	1.86 (0.029)	0.04 (0.017)	4.10 (0.008)	−6.79 (0.028)
A−B+	−7.83 (0.018)	−3.48 (0.011)	1.74 (0.018)	−1.85 (0.014)	−0.04 (0.014)	−4.10 (0.004)	6.82 (0.016)
B+C−	−0.23 (0.031)	3.71 (0.020)	2.70 (0.015)	−4.46 (0.041)	−1.28 (0.007)	−2.80 (0.020)	4.22 (0.031)
B−C+	0.21 (0.043)	−3.70 (0.018)	−2.69 (0.024)	4.45 (0.035)	1.28 (0.010)	2.80 (0.015)	−4.23 (0.037)

sider the PF feature in the A+B− condition, which produced an average connection weight of 3.47. In the complementary A−B+ condition, this feature had a weight of nearly the same size, but of opposite sign (i.e., −3.48). A similar pattern is evident for all of the features, in either the AB or the BC discriminations.

IV. DISCUSSION

Previously, Charrier *et al.* (2005) studied how black-capped chickadees respond to artificial notes that had their entire spectrogram shifted up or down in frequency. They reported a systematic pattern of note misclassifications by these animals. Nickerson *et al.* (2006) replicated this study in artificial neural networks, and found a very strong agreement between the behavior of their networks and the behavior of the birds. There were two limitations of both of these studies. First, by shifting the frequency of the entire note, all of the frequency features were manipulated at the same time. Second, by only shifting note frequency, none of the duration features of the spectrogram were manipulated at all. The purpose of the above-reported simulations was to examine the response of artificial neural networks that had every note feature (frequency and duration) manipulated independently. In the following discussion, we first consider the effects of manipulating the frequency features. We then consider the effects of altering the duration features. Finally, we reflect upon the implications of these simulated results.

A. Frequency manipulations

An examination of Table II reveals several interesting patterns of results concerning the independent manipulation of the frequency features of the notes. The most general of these is that changing any of the frequency features (SF, PF, EF, or Fmax) produced changes in the responses in at least some of the trained perceptrons. That is, none of the columns of Table II for these features are empty.

Of more interest is the fact that not all of these features were of equal importance. In particular, Table II demonstrates that ten different effects were observed when PF was manipulated, and eight different effects were detected when Fmax was changed. In contrast, altering SF and EF produced only four observable effects apiece.

Of most interest is the fact that the effect of manipulating a frequency feature was highly context dependent. That is, whether a manipulation had an effect, and the nature of the effect that was produced, depended critically upon the type of discrimination learning that a perceptron had been given prior to being presented the manipulated notes. For example, consider SF. When networks had undergone AB discrimination learning, decreases in SF made A notes more B-like, and increases in SF made B notes more A-like. However, when networks had undergone BC discrimination learning, manipulations of SF had no effect on network responses at all.

The context dependence of PF manipulations is much more striking. First, consider the context of AB discrimination learning. Decreases in PF made both A notes and B notes more B-like in this context, while increases in PF made

B notes more A-like. In the context of BC discrimination learning, decreases in PF made B notes more C-like, while increases in PF made C notes more B-like. Notice the striking context dependence on responses to B notes: In one context, decreases in PF produce network responses that indicate that B notes are more B-like, but in the other context these same decreases generate network responses that indicate that B notes are less B-like.

The context dependence of EF manipulations is also evident. In the AB context, increases in EF make A notes more B-like, but no other effects are evident. In particular, manipulating EF does not affect B notes in this context. However, in the BC context, decreases in EF make B notes more C-like.

Finally, consider the pattern of results produced by Fmax manipulations. In the AB context, increases in Fmax made A notes more B-like, and decreases in Fmax made B notes more A-like. In the BC context, increases in Fmax made C notes more B-like while decreases in Fmax made B notes more C-like. These results are the least context-dependent of the frequency manipulations, in the sense that there was a qualitatively consistent pattern of effects on B notes across the two learning contexts.

How do these simulated frequency results compare to results from black-capped chickadees? In their extensive bioacoustic analysis of the chick-a-dee call, Charrier *et al.* (2004) examined the extent to which each summary feature was important for determining a particular, and unmodified, note type (A, B, or C). They did this by computing the potential for note-type coding (PNTC) for each feature. The PNTC is an adaptation of the method for the potential for individual identity coding in animal vocalizations (Charrier *et al.*, 2004; Robisson *et al.*, 1993), and is computed for a feature by taking the ratio of its between-note-type variation to its within-note-type variation. A high PNTC value for an acoustic feature suggests that the feature is crucial for classifying the note as belonging to one of the three note types (A, B, and C). Charrier *et al.* (2004) found that SF had a particularly high PNTC value (2.7), and that EF had a moderately high PNTC value (1.6). They also found that PF and Fmax had PNTC values near unity (1.06 and 1.03, respectively), indicating that these features were not crucial for distinguishing types of notes. All of these results are at odds with the above-reported simulation results. However, one factor that is absent from PNTC calculations is a learning context. The simulation results strongly suggest that which features are important for note discrimination depends upon which two notes are being discriminated.

B. Duration manipulations

The general pattern of results revealed for frequency features can also be seen in Table II for the three different duration features of ascending duration (AD), descending duration (DD), and total duration (TD). First, manipulations of all three features produced changes in behavior in some of the perceptrons. Second, some of the duration features appeared to be more important than others: AD and TD both

reveal eight effects in Table II, while DD only reveals two. Third, the effects of manipulating a note's features once again appear to be context dependent.

The most striking example of context dependence is evident in manipulations of DD. In the AB context, altering this feature does not generate any changes in perceptron behavior. In the BC context, increases in DD cause B notes to become more C-like, but no other effects are evident.

Manipulations of the other two duration variables appear to be less context dependent, in the sense that these manipulations appear to have similar effects on B notes in both discrimination learning contexts. In the AB context, increases in AD cause B notes to become more A-like, while decreases in AD cause A notes to become more B-like. In the BC context, increases in AD cause B notes to become more C-like, while decreases in AD cause C notes to become more B-like. The identical pattern of results is revealed in the manipulations of TD in the two contexts.

In terms of relating these results to previous studies, Charrier *et al.* (2004) found that TD had a PNTC value of 1.6, AD had a PNTC value of 1.5, and DD had a PNTC value of 1.3. The fact that DD had a lower PNTC value than did the other two duration features is consistent with our finding that DD did not play as important a role for the perceptrons as did the other two features. However, as with the frequency effects, discussed earlier, these PNTC values did not take context dependence into account.

C. Implications

What are the implications of the simulation results that we have reported? One implication comes from a standard empiricist approach to evaluating models: A model is only as good as the number of new experiments that it generates. Our simulations have revealed some general regularities that need to be explored using birds as subjects. First, the perceptrons reveal that their behavior can be changed by altering individual spectrogram features. Second, the perceptrons suggest that some of these manipulations are more powerful—or at least more likely to produce effects—than are others. Third, the perceptrons indicate that the effect of manipulating a particular feature is strongly context dependent.

Here we do not claim that black-capped chickadees can independently modify the features in their call notes in exactly the same manner that we have in this report. In fact, there is a good likelihood that they may not be able to manipulate individual features in their notes without also affecting other features in the process due to anatomical and physiological constraints on the vocal apparatus (reviewed in Podos and Nowicki, 2004; Suthers, 2004). That being said, there is natural variation in all of the acoustic features in normal chickadee call notes that we manipulated here to produce modified stimuli for our networks (Charrier *et al.*, 2004). Therefore, in spite of the fact that birds may not necessarily be able to produce notes of the type that we are using to test network performance they do produce notes in a variable manner. Further, what we are attempting to do here is to determine which feature (or features), observed to vary in nature during normal vocal production, is (or are) the most important for controlling note type discrimination and categorization. That being said, further studies on anatomical and physiological performance limits on vocal production in chickadees would be an interesting avenue for future research.

The second implication reveals another important contribution that simulation studies can make to empirical research. Animal research is both costly and time consuming. It would be highly advantageous to have *a priori* knowledge about what effects are likely to be revealed in a study that uses animals as subjects. The above-reported simulations provided hypotheses about the potential effects of a large number of feature manipulations—recall that the test set consisted of 4200 different notes. To the extent that the perceptrons provide an applicable model of note discrimination in birds (a fact that was previously established by Nickerson *et al.*, 2006), they have permitted a search of a staggeringly large set of manipulations in an attempt to identify experimental variables that are likely to produce results. Our current work involves testing the hypotheses generated by these simulations using real birds as subjects.

The third implication, that context plays a role in learning and generalization by the perceptrons, places this research into a broader framework of contextual effects on learning and cognition. As noted in Sec. I, context can overshadow CSs, and cause previously extinguished behavior to reemerge. Moreover, and important for the current research, songbirds performing auditory discriminations appear to be very sensitive to the context during generalization tests with *pure tonal* stimuli (Lee *et al.*, 2006), to the extent that without a familiar training context, birds fail to show generalization to novel stimuli. Here, modeling *call note* discriminations using perceptrons, we show a similar and striking effect of training context on acquisition, with significant differences among the different discrimination tasks, and of generalization, with differential responding to transfer notes contingent upon the discrimination context of the perceptrons. Whether this contextual effect is observed in birds performing these same discriminations presents an exciting future avenue of research.

Bouton, M. E. (**2004**). "Context and behavioral processes in extinction," Learn. Memory **11**, 485–495.

Charrier, I., Bloomfield, L. L., and Sturdy, C. B. (**2004**). "Note types and coding in parid vocalizations I. The chick-a-dee call of the black-capped chickadee (*Poecile atricapilla*)," Can. J. Zool. **82**, 769–779.

Charrier, I., Lee, T. T. Y., Bloomfield, L. L., and Sturdy, C. B. (**2005**). "Acoustic mechanisms of note-type perception in black-capped chickadee (*Poecile atricapillus*) calls," J. Comp. Psychol. **119**, 371–380.

Charrier, I., and Sturdy, C. B. (**2005**). "Call-based species recognition in black-capped chickadees," Behav. Processes **70**, 271–281.

Dawson, M. R. W. (**2004**). *Minds And Machines: Connectionism And Psychological Modeling* (Blackwell, Malden, MA).

Dawson, M. R. W. (**2005**). *Connectionism: A Hands-On Approach*, 1st ed. (Blackwell, Oxford, UK).

Dawson, M. R. W., Bloomfield, L. L., Charrier, I., and Sturdy, C. B. (**2006**). "Statistical classification of black-capped (*Poecile atricapillus*) and mountain chickadee (*Poecile gambeli*) call notes," J. Comp. Psychol. **120**, 147–153.

Dawson, M. R. W., Charrier, I., Sturdy, C. B. (**2006**). "Using an artificial neural network to classify black-capped chickadee (*Poecile atricapillus*) call note types," J. Acoust. Soc. Am. **119**, 3161.

Dawson, M. R. W., and Spetch, M. L. (**2005**). "Traditional perceptrons do not produce the overexpectation effect," Neur. Info. Proc. — Letters and Reviews, found at http://nip-lr.info/V07N01/V07N01.htm, **7**, 11–17.

Ficken, M. S., Ficken, R. W., and Witkin, S. R. (**1978**). "Vocal repertoire of black-capped chickadee," Auk **95**, 34–48.

Ficken, M. S., Hailman, E. D., and Hailman, J. P. (**1994**). "The chick-a-dee call system of the Mexican chickadee," Condor **96**, 70–82.

Freeberg, T. M., and Lucas, J. R. (**2002**). "Receivers respond differently to chick-a-dee calls varying in note composition in Carolina chickadees," Anim. Behav. **63**, 837–845.

Hailman, J. P. (**1985**). "Ethology, zoosemiotic and sociobiology," Am. Zool. **25**, 695–705.

Hailman, J. P., and Ficken, M. S. (**1986**). "Combinatorial animal communication with computable cyntax—Chick-a-dee calling qualifies as language by structural linguistics," Anim. Behav. **34**, 1899–1901.

Lee, T. T. Y., Charrier, I., Bloomfield, L. L., Weisman, R. G., and Sturdy, C.

B. (**2006**). "Frequency-range discriminations and absolute pitch in black-capped chickadees (*Poecile atricapillus*), mountain chickadees (*Poecile gambeli*), and zebra finches (*Taeniopygia guttata*)," J. Comp. Psychol. **120**, 217–228.

Mammen, D. L., and Nowicki, S. (**1981**). "Individual-differences and within-flock convergence in chickadee calls" Behav. Ecol. Sociobiol. **9**, 179–186.

Miller, R. R. (**2006**). "Challenges facing contemporary associative approaches to acquired behavior," Comparative Cognition & Behavior Reviews, **1**, 77–93. Retrieved from http://psyc.queensu.ca/ccbr/index.html

Minsky, M., and Papert, S. (**1988**). *Perceptrons*, 3rd ed. (MIT, Cambridge, MA).

Nickerson, C. M., Bloomfield, L. L., Dawson, M. R. W., and Sturdy, C. B. (**2006**). "Artificial neural networks that discriminate notes from the chick-a-dee call of *Poecile atricapillus*: The effect of pitch transformations," J. Acoust. Soc. Am. **120**, 1111–1117.

Nowicki, S. (**1983**). "Flock-specific recognition of chickadee calls," Behav. Ecol. Sociobiol. **12**, 317–320.

Nowicki, S., and Nelson, D. A. (**1990**). "Defining natural categories in acoustic-signals—Comparison of 3 methods applied to chick-a-dee call notes," Ethology **86**, 89–101.

Podos, J., and Nowicki, S. (**2004**). "Performance limits on birdsong," in *Nature's Music: The Science of Birdsong*, edited by P. Marler and H. Slabbekoorn (Elsevier, New York), pp. 318–342.

Robisson, P., Aubin, T., and Bremond, J. C. (**1993**). "Individuality in the voice of the emperor penguin Aptenodytes-Forsteri—Adaptation to a noisy environment," Ethology **94**, 279–290.

Rosenblatt, F. (**1962**). *Principles of Neurodynamics* (Spartan, Washington, DC).

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (**1986**). "Learning representations by back-propagating errors," Nature (London) **323**, 533–536.

Sturdy, C. B., Phillmore, L. S., and Weisman, R. G. (**2000**). "Call-note discriminations in black-capped chickadees (Poecile atricapillus)," J. Comp. Psychol. **114**, 357–364.

Suthers, R. A. (**2004**). "How birds sing and why it matters," in *Nature's Music: The Science of Birdsong*, edited by P. Marler and H. Slabbekoorn (Elsevier, New York), pp. 272–295.

Templeton, C. N., Greene, E., and Davis, K. (**2005**). "Allometry of alarm calls: Black-capped chickadees encode information about predator size," Science **308**, 1934–1937.

Yaremchuk, V., Willson, L. R., Spetch, M. L., and Dawson, M. R. W. (**2005**). "The implications of null patterns and output unit activation functions on simulation studies of learning: A case study of patterning," Learn Motiv **36**, 88–103.

Representing multiple discrimination cues in a computational model of the bottlenose dolphin auditory system

Brian K. Branstetter[a)]
Marine Mammal Research Program, Hawaii Institute of Marine of Marine Biology, P.O. Box 11066, Kailua, Hawaii 96734

Eduardo Mercado III
Department of Psychology, Park Hall, University at Buffalo, SUNY, Buffalo, New York 14260

Whitlow L. Au
Marine Mammal Research Program, Hawaii Institute of Marine of Marine Biology, P.O. Box 11066, Kailua, Hawaii 96734

A computational model of the dolphin auditory system was developed to describe how multiple discrimination cues may be represented and employed during echolocation discrimination tasks. The model consisted of a bank of gammatone filters followed by half-wave rectification and low pass filtering. The output of the model resembles a spectrogram; however, the model reflects temporal and spectral resolving properties of the dolphin auditory system. Model outputs were organized to represent discrimination cues related to spectral, temporal and intensity information. Two empirical experiments, a phase discrimination experiment [Johnson *et al.*, Animal Sonar Processes and Performance (Plenum, New York, 1988)] and a cylinder wall thickness discrimination tasks [Au and Pawolski, J. Comp. Physiol. A **170**, 41–47 (1992)] were then simulated. Model performance was compared to dolphin performance. Although multiple discrimination cues were potentially available to the dolphin, simulation results suggest temporal information was used in the former experiment and spectral information in the latter. This model's representation of sound provides a more accurate approximation to what the dolphin may be hearing compared to conventional spectrograms, time-amplitude, or spectral representations. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2772214]

I. INTRODUCTION

Dolphins have an acute ability to echoically discriminate between targets that vary along the dimensions of size (Ayrapet'yants *et al.*, 1969; Au and Pawloski, 1992), material composition (Nachtigall, 1980), and shape (Helweg *et al.*, 1996; Pack *et al.*, 2002). During echolocation, dolphins typically emit a series of "clicks" where the inter-click interval is sufficient for the dolphin to receive an echo before emitting a consecutive click. Clicks, or incident signals, can be described as broadband transients with peak frequencies between 40 and 140 kHz, durations as short as 40 μs, and peak-to-peak sound pressure levels often exceeding 220 dB re : 1 μPa (Au, 1980). The signal is projected forward in a tight beam, elevated about 5° above the rostrum, and having a 3 dB beamwidth of approximately 10° (Au, 1980).

Figure 1 illustrates a single on-axis dolphin click in the time domain, and its power spectral density function. Dolphin clicks are highly directional and can display considerable spectral-temporal, off-axis variation (Au, 1980; Au *et al.*, 1986). Such graphical representations have been invaluable for describing the physical attributes of dolphin phonations, given that much of the sounds that dolphins produce

are well out of the human frequency sensitivity range. Because dolphins produce and hear high frequency broadband sounds (Johnson, 1967; Au, 1980), it is likely that these sounds are biologically significant (Lammers *et al.*, 2003). However, the graphical representations typically employed may not accurately reflect how a dolphin represents or perceives clicks and their echoes. For example, Lammers *et al.* (2003) demonstrated that a conventional spectrogram representation of spinner dolphin (*Stenella longirostris*) whistles may fail to highlight meaningful perceptual properties. Higher frequency whistle harmonics appear lower in amplitude than the fundamental frequency when a spectrogram displays sound pressure level [Lammers *et al.*, 2003; Fig. 7A]. However, when a correction factor is applied that takes into account the dolphin's frequency hearing sensitivity (sensation level), the first and second harmonics are likely to be "perceived" as being louder than the fundamental [Lammers *et al.*, 2003; Fig. 7B]. Because the higher frequency harmonics are relatively more directional, listening dolphins could use these harmonics to estimate the direction other dolphins are oriented, and thus, the direction they are traveling. Lammers *et al.* (2003) suggested that the information derived from harmonics may be important for maintaining group cohesion in a visually restricted environment, and their importance may have gone unrecognized using conventional graphical representations.

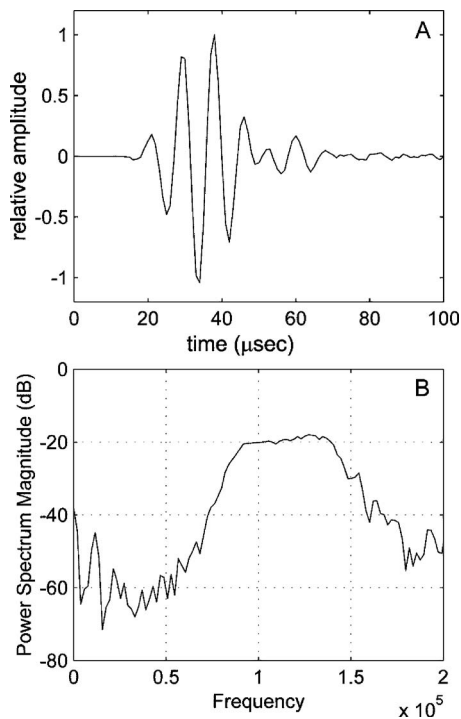
a)Electronic mail: branstet@hawaii.edu

FIG. 1. (A) Time amplitude representation of a bottlenose dolphin "click" recorded directly in front of the dolphin (on-axis) with (B) the power spectrum magnitude of the click.

Similarly, the importance of specific echo features during an echolocation task could also go unrecognized if our representations fail to include perceptual qualities. The first goal of this study was to develop a biologically based representation of sound that reflects processing stages within the dolphin auditory system. These processing stages place limits on frequency and temporal resolution, both of which have been hypothesized to play important roles in echoic discrimination abilities (Au, 1988; Au and Nachtigall, 1995). The second goal was to use the perceptual model to simulate dolphin auditory discrimination capabilities in two experiments: (1) discrimination of cylinder wall thickness differences (Au and Pawloski, 1992), and (2) sensitivity to differences in signal phase (Johnson et al., 1988). These two experiments were chosen because the dolphins were likely to employ frequency and/or temporal information to perform the discrimination. In addition, the current model builds on previous modeling efforts (Au, 1994; Au et al., 1995; Johnson et al., 1988) that have focused on these two experiments. Unlike the prior modeling efforts, the current model assumes time, frequency, and energy domain information (potential discrimination cues) are available simultaneously, are processed by separate modules, and that the relative saliency of each domain will determine which cue the dolphin likely uses for discrimination.

Time-frequency resolution. A three-dimensional object can reflect sound from both external and internal surfaces separated in space. These multiple reflections, within a single echo, are often called echo highlights and produce amplitude modulation in the envelope of the temporal wave form, as well as peaks and valleys in the echo spectrum. Dolphins can potentially use both envelope and spectral profile informa-

tion for discriminating between objects (Johnson et al., 1988; Au and Pawloski, 1992). The dolphin's ability to resolve time-frequency information will be limited by the resolving qualities of at least two distinct auditory mechanisms: (1) the auditory periphery (Au and Moore, 1990; Lemonds, 1999; Finneran et al., 2002) and (2) central auditory mechanisms (Viemeister, 1979). The auditory periphery (i.e., basilar membrane response) is often modeled as a series of continuously overlapping bandpass filters (Fletcher, 1940) where the filter bandwidth is proportional to the center frequency (Au and Moore, 1990). Spectral resolution is then directly proportional to filter width, with the trade-off of inversely proportional temporal resolution. This results in good frequency resolution at the lower end of the sound spectrum, and good temporal resolution at the higher end.

After transduction and filtering at the basilar membrane, temporal resolution will be limited by the refractory period of the eighth nerve and central auditory areas. Because neurons have refractory periods (whereas hair cells on the basilar membrane have graded potentials limited only by the speed of ionic displacement) temporal processing at the eighth nerve is relatively sluggish.

II. COMPUTATIONAL MODEL OF THE DOLPHIN AUDITORY SYSTEM

A. Auditory peripheral filters

The relationship between the center frequency of an auditory filter and the filter bandwidth can be described by

$$Q = f_0/\Delta f, \tag{1}$$

where f_0 is the center frequency of the filter and Δf is the filter bandwidth. Bandwidth measurements for *Tursiops truncatus* using broadband maskers (critical ratio: Johnson, 1968) and band limited maskers (critical band: Au and Moore, 1990) have been measured at 12.3 and 2.2, respectively.

Precise auditory peripheral filter shapes were measured for *T. truncatus* employing a tone-detection in notched-noise paradigm (Lemonds, 1999; Finneran et al., 2002). Filter shapes were described by a two-parameter rounded exponential (roex) function defined by

$$W(g) = (1 - r)(1 + pg)\exp(-pg) + r, \tag{2}$$

where p and r determine the slope and shape of the filter and g is

$$g = \frac{|f - f_c|}{f_c}. \tag{3}$$

The parameter f is a frequency array in hertz, and f_c is the center frequency of the filter. Although the roex filter is commonly used to describe spectral filtering at the auditory periphery, models attempting to more accurately simulate temporal features of auditory processing (such as phase locking characteristics) have often relied on the gammatone filter (Patterson et al., 1992). Gammatone filters appear to model a variety of human psychophysical and physiological data well, and they have a well-defined impulse response that is sensitive to phase information (Patterson et al., 1992). In

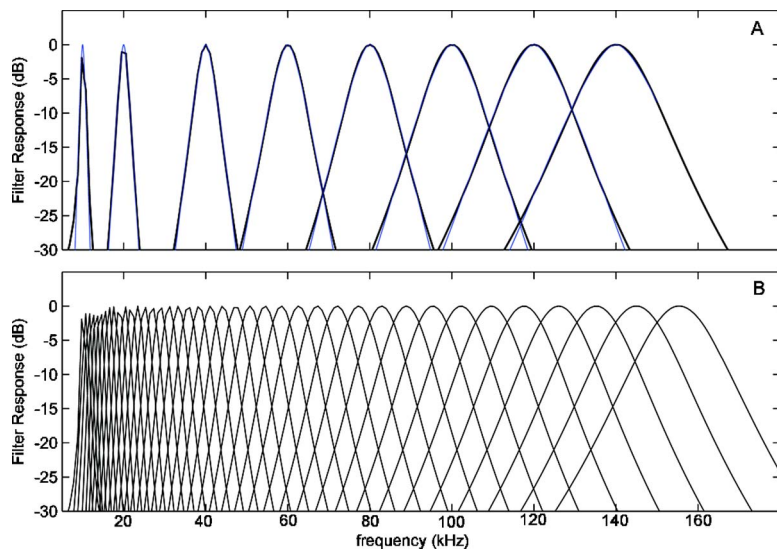Branstetter *et al.*: Computational model of dolphin discrimination cues

FIG. 2. (Color online) (A) roex filters (Lemonds, 1999) superimposed with gammatone filters from the current model. The filters are almost identical to −30 dB. (B) frequency response of 50 gammatone filters spaced between 11 and 150 kHz.

addition, the Fourier transform of the gammatone filter is nearly identical to the roex filter.

The impulse response of the gammatone filter defined by Patterson (1994) for $t > 0$ is

$$g_t(t) = at^{(n-i)}\exp(-2\pi bt)\cos(2\pi f_c t - \phi), \tag{4}$$

where f_c is the center frequency of a channel in hertz, ϕ is the starting phase, and a, b, n, are parameters determining ramping and duration of the impulse response and thus filter shape and width in the spectral domain (Slaney, 1993).

The current model employs a gammatone implementation (Slaney, 1993) originally fit to human data. Modifications were made to the current form to reflect dolphin broadband listening and spectral resolution. The parameter b is related to the equivalent rectangular bandwidth (ERB) by $b = k^*\mathrm{ERB}(f_c)$, where k is a constant equal to 1.019 (Patterson et al., 1992). ERBs are psychophysical measures of auditory filters derived from masking experiments (Fletcher, 1940; Au and Moore, 1990; Lemonds, 1999; Finneran et al., 2002). For current modeling purposes, a simplified form for the ERB was used

$$\mathrm{ERB}(f_c) = \frac{f_c}{Q} + \mathrm{MinBW}, \tag{5}$$

where f_c and Q were previously defined. MinBW is the minimum bandwidth for low frequency channels and in humans is estimated at 24.7 (Glasberg and Moore, 1990). Employing the human value of 24.7 is acceptable because for dolphin high frequency listening (with large ERBs), adding the constant of 24.7 becomes negligible.

When $Q = 11.3$, the current gammatone filter bank produces excellent fits to two-parameter roex filters ($r = 0.0041, p = 29.49$) derived from dolphin empirical data (Lemonds, 1999). Figure 2(a) illustrates roex filters (Lemonds, 1999) and gammatone filters (current model) plotted against each other. The ten filters are spaced between 10 and 140 kHz and are nearly identical down to −30 dB. The current Q value is within the range estimated for the bottlenose dolphin (Johnson, 1968; Au and Moore, 1990). The number of filters and frequency range of the filters are free param-

eters in the model. Filters are spaced apart so there is an equal percentage of overlap between each filter's ERB. Figure 2(b) illustrates the frequency response of 50 filters spaced between 11 and 156 kHz. Figure 3 illustrates the impulse response of the same 50 filters.

B. Hair cell model

The output of each gammatone filter is then, in parallel, half-wave rectified as described by

$$f_{\mathrm{rect}}(t) = (f(t) + \sqrt{f(t)^2})/2. \tag{6}$$

Nonlinear rectifiers (such as half wave rectification) mimic the afferent activity of an inner hair cell on the basilar membrane (Regan, 1994). Transduction occurs only when the stereo cilia of inner hair cells are unidirectionally displaced toward the stria vascularis (Hudspeth, 1983). Inner hair cells, therefore, produce a graded potential per cycle in a direct coupling with the physical stimulus.
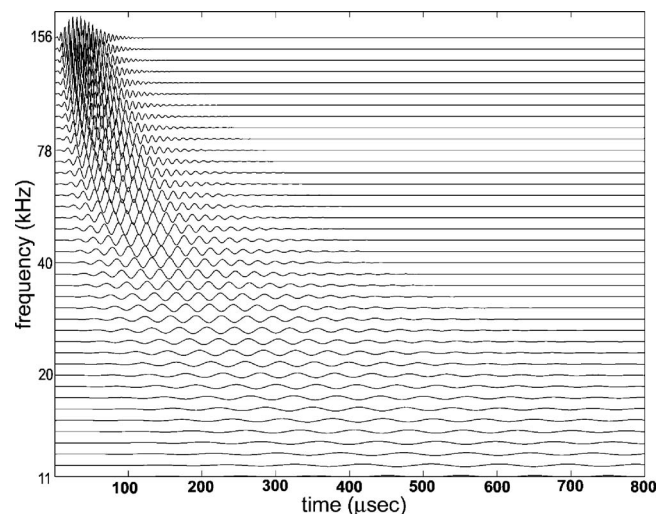

FIG. 3. Impulse response of the gammatone filter bank. The output of each filter is the relative time-amplitude response and the y axis labels correspond to the center frequency of the filter.
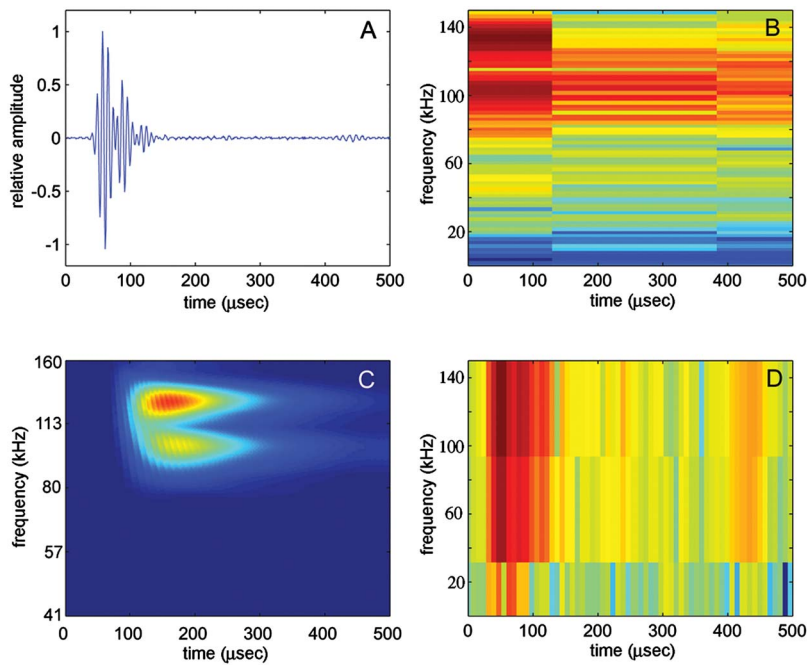
FIG. 4. (Color online) Four different representations of the same echo. (A) Echo in the time domain. (B) Spectrogram with 512 point Hanning window. (C) Tursiogram. (D) Spectrogram with 16 point Hanning window.

C. Leaky integrator low-pass filter

The sluggishness of the eighth nerve and higher auditory areas can be modeled by a low-pass filter defined in the time domain by the exponential decay window

$$h(t) = ke^{-t/\tau}, \tag{7}$$

where k is a constant scaling factor, t is units of time, and $\tau = 264$ μs, which is the dolphin critical interval (Vel'min and Dubrovskii, 1976; Moore *et al.*, 1984) or integration time (Au and Moore, 1988). The output of the low pass filter is the convolution:

$$g(t) = \int_0^T f_{\text{rect}}(t)^* h(t - \tau) dt. \tag{8}$$

The $h(t)$ is often referred to as a "leaky integrator" and the output $g(t)$ resembles the envelope of the temporal wave form calculated by the Hilbert transform. However, unlike the Hilbert transform, the leaky integrator output captures the phase locking capability (for low frequency sounds) and the envelope following response of the hair cell/auditory nerve complex. Furthermore, leaky integrator output incorporates the low-pass filtering characteristics of the dolphin auditory neurons. In the frequency domain, $G(S)$ is calculated by the equation $G(S) = F_{\text{rect}}(S)^* H(S)$, where $F_{\text{rect}}(S)$ and $H(S)$ are the Fourier transforms of $f_{\text{rect}}(t)$ and $h(t)$, respectively. Because multiplication is computationally more efficient than evaluating the convolution integral in Eq. (8), all model computations were conducted in the spectral domain. Visual representations of the output $g(t)$ resemble spectrograms, however, spectral and temporal resolution of the bottlenose dolphin define the output. Hence, the output $g(t)$ is called the tursiogram. In summary, sounds presented to the model were bandpass filtered by a bank of gammatone filters; the output of each filter was then half-wave rectified, and low-pass filtered.

Time-frequency representations (spectrograms) of individual echoes are seldom employed due to an echo's short duration. Figure 4 displays four different representations of an echo from a hollow cylinder sampled at 1 MHz. Figure 4(A) is the time-amplitude representation. Figure 4(B) is a spectrogram with a FFT length of 512, Hanning window length of 512, with 50% overlap. Figure 4(C) is a 200 channel tursiogram. Figure 4(D) is a spectrogram with a FFT length of 16, Hanning window length of 16, with 50% overlap. The spectrograms in Figs. 4(B) and 4(D) demonstrate problems associated with the spectral-temporal trade-off inherent with conventional spectrograms. Figure 4(B) has good spectral resolution at the cost of poor temporal resolution. The opposite holds true for Fig. 4(D), which has good temporal resolution but poor spectral resolution. Neither representation captures the frequency spacing, frequency resolution, or temporal resolution specific to the dolphin auditory system. The tursiogram in Fig. 4(B) is more similar to what a dolphin actually hears than an arbitrary time-frequency representation.

III. REPRESENTING MULTIPLE DISCRIMINATION CUES

Human listeners can discriminate short duration sounds based on perceptual differences related to spectral profile (pitch), intensity (loudness), and the temporal envelope (roughness, timbre, and time separation pitch). Furthermore, listeners can preferentially select a specific cue when multiple cues are simultaneously available (Southworth and Berg, 1995). The ability to attend to different simultaneously available cues is further supported by the presence of auditory cortical fields with differential sensitivities to different stimulus attributes (Merzenich and Schreiner, 1992). Evidence suggests dolphins can also discriminate between sounds differing in cues related to frequency (Herman and Arbeit, 1972), intensity (Evans, 1973), and the temporal en-

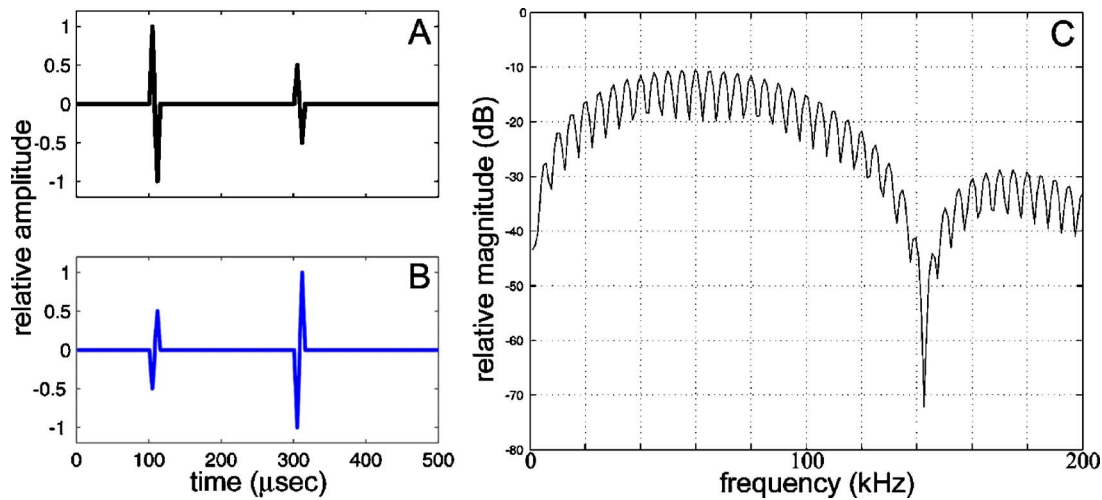Branstetter *et al.*: Computational model of dolphin discrimination cues

FIG. 5. (Color online) Click stimuli [(A) and (B)] and their identical spectra (C) superimposed. Note that the two clicks have identical spectra.

velope (Helweg *et al.*, 1996). Zorikov and Dubrovski (2003) demonstrated that when multiple discrimination cues were simultaneously available, a dolphin preferentially selected individual cues in a hierarchical fashion. For short duration stimuli (<264 μs), the envelope or "macrostructure" of the power spectrum was the preferred cue, followed by the fine structure of the spectral profile (i.e., spectral peaks and valleys), with energy the least likely cue to be employed. Although the authors suggest the discrimination cues may be preferentially organized, the dolphin's cue preference may be the result of the availability or saliency of each cue. Cue availability will depend on at least the acoustic dimensions (temporal, spectral, intensity) of the stimulus and the dolphin's ability to resolve each of these dimensions.

The tursiogram was extended to organize received sounds into three different perceptual representations (temporal envelope, spectral profile, and intensity). The models were then tested against each other in simulations of two discrimination tasks previously performed by dolphins to determine which cue was most likely employed by the dolphin.

A. Envelope detection model

Viemeister and Plack, 1993 suggested that information from independent peripheral filters may be pooled to form a "reasonably faithful recovery of the envelope." The envelope detection model (EDM) does just that. The EDM has four stages. The first three stages are the tursiogram (gammatone filtration, half-wave rectification, low pass filtering) followed by across channel summation. Across channel summation is used to produce a single representation of the hypothetically "perceived" envelope ENV(t) and is calculated by

$$\text{ENV}(t) = \sum_{i=1}^{N} g_i(t),\tag{9}$$

where $g(t)$ is the time-domain [inverse Fourier transform of $G(S)$] output from the low pass filter (from the tursiogram) of the ith frequency channel. The EDM resembles the Hilbert transform envelope, however, the dolphin's temporal resolution is accounted for.

B. Spectral profile model

The majority of dolphin biomimetic sonar models base decisions on the spectral profile of the echo (Roitblat *et al.*, 1993; Au, 1994; Au *et al.*, 1995; Roitblat *et al.*, 1996). The current spectral profile model (SPM) has four stages: gammatone filtering, half-wave rectification, low pass filtering (from tursiogram), and within channel summation. Within channel summation assumes the energy within each channel is summed and can be computed by

$$E_i = \sum_{t=1}^{N} g_i(t),\tag{10}$$

where i is the ith frequency channel.

The output of the SPM resembles the power spectral density function of the original signal with the dolphin's temporal and spectral resolution accounted for.

C. Intensity model

Level cues due to the intensity of a stimulus can be represented by applying both within channel summation (SPM) and across channel summation (EDM). The result is a single value representing the hypothetically perceived intensity (I) of the stimulus,

$$I_{it} = \sum_{t=1}^{N} \sum_{i=1}^{N} g_i(t).\tag{11}$$

IV. SIMULATIONS OF DISCRIMINATION BY ECHOLOCATING DOLPHINS

A. Temporal order discrimination

In a phase order discrimination experiment (Johnson *et al.*, 1988), a dolphin was trained to discriminate between two click train stimuli: (1) a large click preceding a small click [Fig. 5(A)]; and (2) a small click preceding a large click [Fig. 5(B)]. Each click was a single cycle from a 60 kHz sinusoid and the time separation between the clicks was 200 μs. Each click train was composed of 30 echoes. A dolphin was able
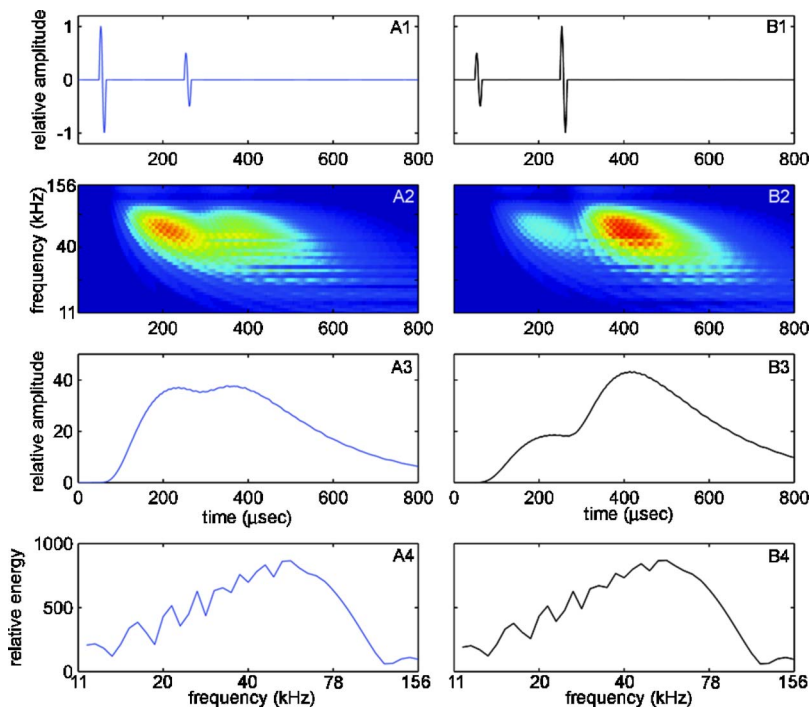
FIG. 6. (Color online) (A1), (B1) the S+ and S− stimuli respectively. (A2), (B2) Tursiograms from (A1) and (B1). (A3), (B3) EDM outputs of (A1) and (B1). (A4), (B4) The corresponding outputs of the spectral model.

to discriminate between the two stimuli with 75% accuracy despite the fact that the stimuli have identical long-term spectra [Fig. 5(C)], and that the time separation between clicks was smaller than the dolphin critical interval of 264 μs. The authors applied a 300 μs chi-square analysis window to a short-time spectral analysis (spectrogram) to mimic the dolphin's temporal resolution. Although temporal order differences between the spectrogram representations emerged, no attempt was made to incorporate spectral resolution of the dolphin. The current modeling effort incorporates both temporal and spectral limitations of the bottlenose dolphin auditory system.

The phase order discrimination experiment was simulated with the three above-described models using the MATLAB 6.0 programming environment. The goal was to pit the models against each other to determine which model would perform best for this specific task. The cue associated with the best performing model would then be the likely cue the dolphin may have employed during this task. The gammatone filter bank was composed of 40 channels whose center frequencies were spaced between 11 and 156 kHz. Each model was required to perform a two-alternative, match-to-sample task, with the sample and the target stimuli always being the stimulus with the large amplitude click preceding the small amplitude click (S+); the nontarget stimulus was the stimulus with the small amplitude click preceding the large amplitude click (S−). A total of 400 comparisons were made for each model, 200 between the sample and the S+ stimulus, and 200 between the sample and the S− stimulus.

All model decisions were determined by calculating the Euclidean distance between the sample and each alternative. The Euclidean distance measure can be defined as

$$d = |x - y| = \sqrt{\sum_{i-1}^{n} |x_i - y_i|^2}, \qquad (12)$$

where x is the sample vector and y is the alternative vector. Vector values were model dependent. For example, values for the SPM were the summated energy from within each frequency channel, whereas values for the EDM were derived from the instantaneous amplitude of the resulting envelope. The alternative that resulted in the smallest Euclidean distance from the sample was selected as the match, and the percentage of correct selections was calculated. Gaussian noise was added to the stimuli to degrade model performance. The noise level was adjusted so that the best model would perform at 75% correct. Figure 6 displays model outputs for the tursiogram, SPM and EDM. Because the intensity model results in a single value, its output was not displayed. Even before noise is added, casual visual inspection of the model outputs suggest that the EDM [Figs. 6(A3) and 6(B3)] will produce the greatest Euclidean distances, and thus best discrimination performance.

1. Results

Simulation based on the EDM yielded superior performance. Internal noise was added and adjusted to degrade EDM model performance to approximately 75%. This internal noise value was then used for the spectral model and the intensity model, both resulting in chance performance (52% and 46%, respectively).

2. Conclusion

The simulations suggest that the most salient information available to the dolphin within this task would originate from the envelope of the temporal wave form. Changes in the temporal envelope for humans are often perceived as
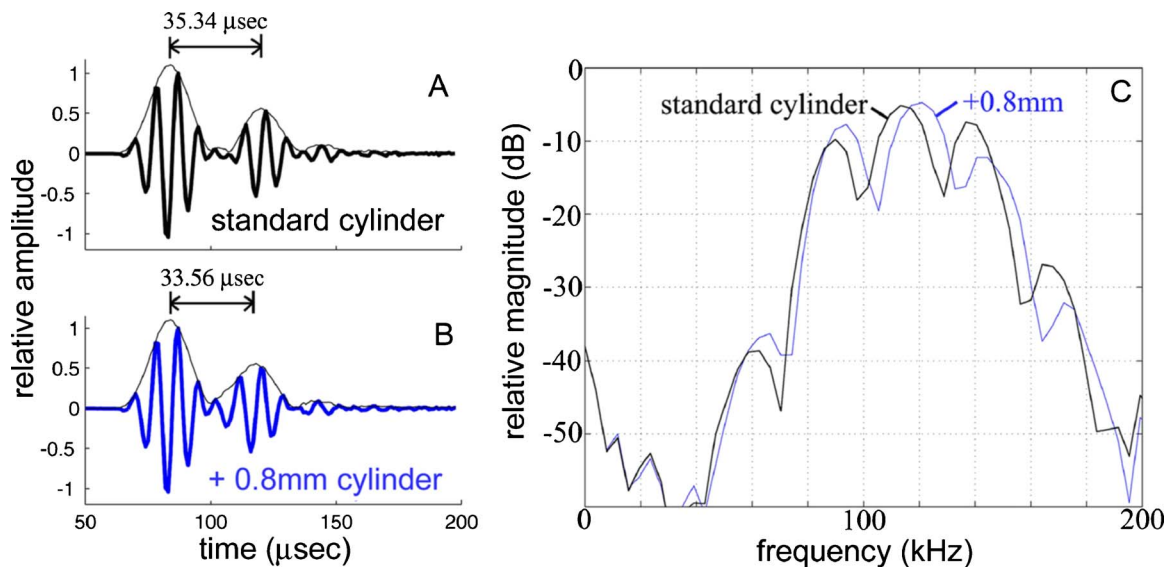
FIG. 7. (Color online) (A) Echo from standard cylinder. (B) Echo from +0.8 mm cylinder. (C) Spectral content of both cylinders.

changes in "roughness" or "time-separation pitch." Au and Martin (1989) reported that human subjects listening to time stretched versions of echoes obtained with simulated dolphin echolocation signals often indicated a time-separation pitch like cue in multihighlight echoes. What the dolphin perceives remains speculative.

B. Simulation of wall thickness discrimination

Au and Pawloski (1992) investigated an echolocating dolphin's ability to discriminate wall thickness differences between hollow cylinders. The dolphin was trained to respond only to a standard cylinder of 6.35 mm wall thickness while the comparison cylinders varied by ±0.2, ±0.3, ±0.4, and ±0.8 mm. All cylinders had identical lengths and outer diameters. Results of the experiment are plotted in Fig. 8. Because the echoes from the cylinders vary in their spectral profiles [Fig. 7(C)] and in their temporal envelopes [Figs. 7(A) and 7(B)], the dolphin could presumably use either of these cues to perform this discrimination task.

Au (1994) simulated the task using a neural network whose inputs were the pattern of energy from a bank of bandpass filters mimicking the frequency selectivity of the auditory periphery. This model was similar in features to the SPM proposed earlier, and provided good fits to dolphin performance. Au et al. (1995) then used a time-frequency model where spectral energy from each channel was evaluated in time increments (temporal windows) related to the reciprocal of filter bandwidths and the dolphin's temporal integration constant. These procedures increased the signal to noise ratio (and thus model performance) by eliminating any noise beyond the temporal windows.

Although all three modeling attempts represent advances in biomimetic modeling, the temporal windows employed make at least three assumptions that can be improved upon: (1) The shapes of the temporal windows are rectangular; (2) temporal information within the windows is not resolvable; and (3) information within each window is evaluated independently (i.e., adjacent temporal information does not effect

each other). Biological auditory systems appear to "smear" information in the time domain rather than partition information into discrete time units (Viemeister, 1979; Bregman, 1990). The exponential decay function (leaky integrator) used in the current model captures this smearing process. Because the dolphin in the cylinder wall thickness discrimination tasks could potentially have based its decisions on either the temporal envelope or spectral information, simulations were performed to evaluate which of these possible cues the dolphin likely used.

The major temporal difference between the echoes from each cylinder is the time separation between the first and second highlights (Au and Pawloski, 1992). For modeling purposes, these stimuli were therefore synthesized using the method developed by Au and Pawloski (1992) using the incident signal reported in Au (1994). Each synthetic echo contained only the first two highlights. The time separation between the first and second highlight ($\Delta\tau$) can be expressed as

$$\Delta\tau = \frac{2\text{th}}{c_1} + \frac{2(\text{o.d.} - 2\text{th})}{c_0}, \tag{13}$$

where th is the cylinder wall thickness, o.d. is the outer diameter equal to 37.85 mm, c_0 and c_1 are the sound velocities in water and the cylinder wall and are equal to 1530 m/s and 5150 m/s, respectively. Once $\Delta\tau$ was calculated for each cylinder, the relative amplitude of the first and second highlights were estimated by averaging ten recorded echoes from the standard cylinder used by Au (1994). The attenuation of the second highlight relative to the first highlight was calculated by dividing the average amplitude of the second highlight by the average amplitude of the first highlight. Two highlight simulated echoes were then synthesized using

$$e(t) = s(t) + k[s(t - \Delta\tau)], \tag{14}$$

where $s(t)$ was the synthetic incident signal from Au (1994), k is the attenuation of the second highlight, and $\Delta\tau$ is the time separation between the first and second highlights.
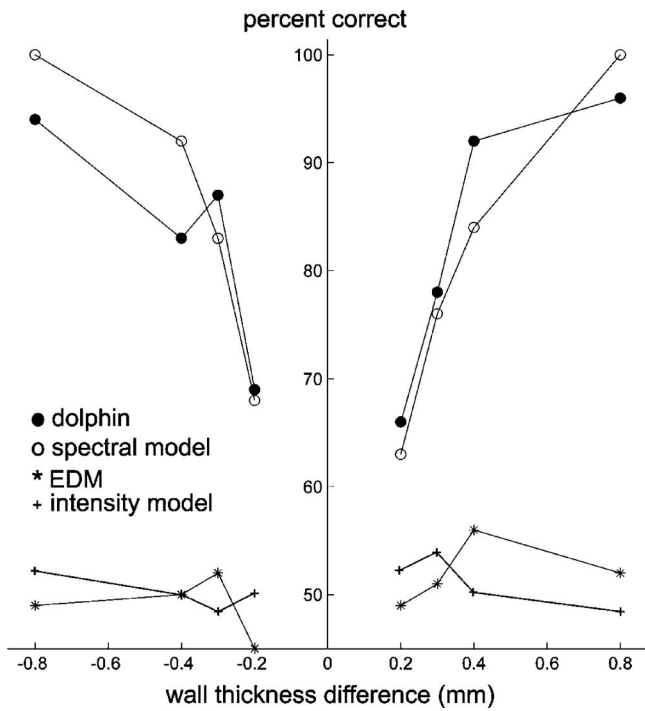
FIG. 8. Results from the cylinder wall thickness discrimination simulation.

Echoes from the standard cylinder (sample) were compared with echoes from two alternatives: the standard (matching alternative) versus a comparison (nonmatching alternative) in a two-alternative, matching-to-sample procedure. Each comparison was performed 200 times. A total of 40 gammatone filters were spaced between 41 and 160 kHz to capture the spectral profile information of the cylinder echoes [Fig. 7(C)]. Euclidean distance measures were used to evaluate similarity and choice of alternative as in the previous simulation.

1. Results

The spectral model outperformed the EDM and intensity models. When noise was added to produce dolphin-like performance, the EDM and intensity model resulted in chance performance (Fig. 8). The output of the EDM produced envelopes that are almost identical [Figs. 9(A3) and 9(B3)]. All highlight information was smeared and unresolvable. However, clear differences in the spectral model outputs [Figs. 9(A4) and 9(B4)] are visible.

2. Conclusion

The spectral model provides the most plausible representation of the information a dolphin might use within this task. The EDM, intensity model, and time-frequency model (Au *et al.*, 1995) are less plausible assuming temporal information within the dolphin integration time (264 μs) is "smeared." These results agree well with simulations from both Au (1994) and Au *et al.* (1995) where spectral information provides the basis for the dolphin's discrimination capabilities. The results also agree with human discrimination abilities reported by DeLong *et al.* (2007). In this study, humans were presented with frequency transposed (downsampled) echoes from the cylinder wall discrimination experiment (Au and Pawloski, 1992). Human listeners performed as well or better than the dolphin. Listeners also reported using a variety of cues (pitch, loudness, duration, timber), however, the vast majority reported using a pitch cue as the basis for discrimination.

V. DISCUSSION

Dolphins are likely to exploit any and all available acoustic information for survival purposes. The type of information available will not only depend on physical prop-
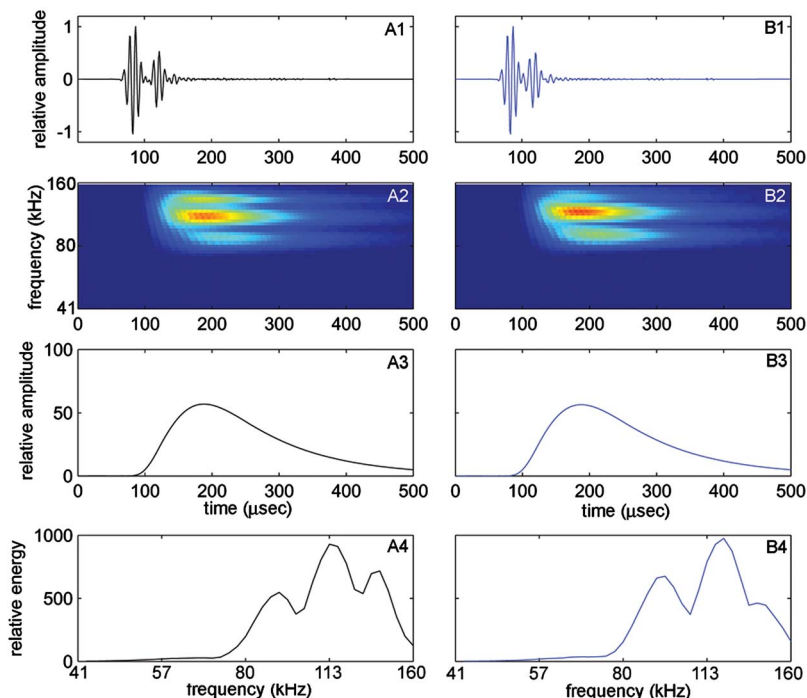

FIG. 9. (Color online) (A1), (A2) Echoes from the standard and +0.8 mm cylinders, respectively. (A2), (B2) Their corresponding tursiograms. (A3), (B3) The corresponding EDM outputs. (A4), (B4) The corresponding outputs of the spectral model.

erties of the sound, but also on how the auditory system represents sound, and the transformations imposed on these representations. The tursiogram provides a time-varying auditory representation reflecting dolphin specific auditory properties not captured in conventional spectrograms, time-amplitude, or spectral representations. The tursiogram provides a more accurate approximation to what the dolphin may be hearing. Although the current model has only been applied to echolocation signals, the model could be useful for examining social sounds (whistles, burst-pulse, tail slapping, etc.) as well. For example, burst-pulse sounds and multiharmonic whistles will likely undergo significant auditory signal processing not captured by conventional representations.

When a dolphin performs an auditory discrimination or identification task, determining what dimension of sound (i.e., loudness, pitch, timbre, spatial location) the dolphin is using to base its decision can be problematic. One method for addressing this problem has been to present humans with the frequency-transposed stimuli and to have them verbally report which discrimination cue they employed (Fish *et al.*, 1976; DeLong *et al.*, 2007). Although this procedure is particularly useful (since few other options are available), caution must be exercised when generalizing human hearing abilities to dolphin hearing. Humans are sensitive to lower frequency sounds and have a longer temporal integration time (DeLong *et al.*, 2007). To accommodate these differences, stimuli initially presented to dolphins are typically transformed by downsampling before being presented to humans. This results in a spectral shift to lower frequencies as well as stretching the sound in the time domain. Thus, dolphin and human subjects are exposed to different stimuli with potentially different cues. Also, dolphins likely organize sound in ways that humans do not. Pack *et al.* (2002) suggest dolphins can perceive the spatial structure or "shape" of object though echolocation. This type of auditory processing would likely require sound to be organized in a high resolution, spatio-topic map of auditory space (Branstetter and Mercado, 2006). If a dolphin made auditory discriminations based on spatial structure, this ability would unlikely be replicated by a human listener.

Simulation results reported here suggest computational modeling can aid in determining what discrimination cues a dolphin may employ. One advantage of the modeling approach is that identical stimuli can be presented to the dolphin and the model. Also, in the current model, all acoustic transformations are based on processes derived from dolphin hearing experiments. Model performance in the cylinder wall thickness discrimination simulation matched closely to dolphin performance, and as with human subjects from DeLong *et al.* (2007), the most salient discrimination cue was the spectral profile or subjective "pitch." In fact, the similarity in cue utilization for both human listeners and the computational model suggests using both approaches concurrently may yield more informative results.

The current model provides only a first approximation to dolphin auditory representations. Additional processing would be required to incorporate processes of sound localization (Renaud and Popper, 1975; Moore *et al.*, 1995; Bran-

stetter and Mercado, 2006; Branstetter *et al.*, 2007) or echoic shape perception (Pack *et al.*, 2002). The tursiogram could serve as a peripheral module for models of these abilities. Improvements to the model could be made by weighting the outputs of the gammatone filter to reflect audiogram sensitivity. Outputs of the model could also be transformed to predict what auditory evoked responses should occur for a given sound. For example, the output of the EDM closely resembles the envelope following response (Dolphin, 1996; Finneran and Houser, 2007) for amplitude modulated stimuli, which corresponds to neural responses from the dolphin's auditory system.

ACKNOWLEDGMENTS

Au, W. W. L. (**1980**). "Echolocation signals of the Atlantic bottlenosed dolphin (*Tursiops truncatus*) in open waters," in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 251–282.

Au, W. W. L. (**1988**). "Detection and recognition models of dolphin sonar systems," in *Animal Sonar Processes and Performance*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 753–768.

Au, W. W. L. and Martin, D. W. (**1989**). "Insights into dolphin sonar discrimination capabilities from human listening experiments," J. Acoust. Soc. Am. **86**, 1662–1670.

Au, W. W. L. and Moore, P. W. B.(**1988**). "The perception of complex echoes by an echolocating dolphin," in *Animal Sonar Processes and Performance*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum Press, New York), pp. 295–299.

Au, W. W. L., Andersen, L. N., Rasmussen, A. R., Roitblat, H. L., and Nachtigall, P. E. (**1995**). "Neural network modeling of a dolphin's sonar discrimination capabilities," J. Acoust. Soc. Am. **98**, 43–50.

Au, W. W. L., and Moore, P. W. B. (**1990**). "Critical ratio and critical bandwidth for the Atlantic bottlenosed dolphin," J. Acoust. Soc. Am. **88**, 1635–1638.

Au, W. W. L., Moore, P. W. B., and Pawloski, D. (**1986**). "Echolocation transmitting beam of the Atlantic bottlenose dolphin," J. Acoust. Soc. Am. **80**, 688–694.

Au, W. W. L., and Nachtigall, P. E. (**1995**). "Artificial neural network modeling of dolphin echolocation," in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden, The Netherlands), pp. 183–199.

Au, W. W. L., and Pawloski, D. A. (**1992**). "Cylinder wall thickness difference discrimination by an echolocating Atlantic bottlenose dolphin," J. Comp. Physiol. **170**, 41–47.

Ayrapet'yants, E. S., Golubkov, A. G., Yershova, I. V., Zhezherin, A. R., Zvorykin, V. N., and Korolev, V. I. (**1969**). "Echolocation differentiation and characteristics of radiated pulses in dolphins," Dokl. Akad. Nauk SSSR **188**, 1197–1199.

Branstetter, B., and Mercado, E. III (**2006**). "Sound localization by cetaceans," J. Comp. Psychol. **19**, 26–61.

Branstetter, B., Mevissen, S. J., Pack, A. A., Herman, L. M., Roberts, S. R., and Carsrud, L. K. (**2007**). "Dolphin (*Tursiops truncatus*) echoic angular discrimination: Effects of object separation and complexity," J. Acoust. Soc. Am. **121**, 626–635.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).

DeLong, C. M., Au, W. W. L., and Stamper, S. A. (**2007**). "Echo features used by human listeners to discriminate among objects that vary in material or wall thickness: Implications for echolocating dolphins," J. Acoust. Soc. Am. **121**, 605–617.

Dolphin, W. F. (**1996**). "Auditory evoked responses to amplitude modulated stimuli consisting of multiple envelope components," J. Comp. Physiol. [A] **179**, 113–121.

Evans, W. E. (**1973**). "Echolocation by marine delphinids and one species of fresh- water dolphin," J. Acoust. Soc. Am. **54**, 191–199.

Finneran, J. J., and Houser, D. S. (**2007**). "Bottlenose dolphin (*Tursiops truncatus*) steady-state evoked responses to multiple simultaneous sinusoidal amplitude modulated tones," J. Acoust. Soc. Am. **121**, 1775–1782.

Finneran, J. J., Schlundt, C. E., Carder, D. A., and Ridgway, S. H. (**2002**). "Auditory filter shapes for the bottlenose dolphin (*Tursiops truncatus*) and the white whale (*Delphinapterus leucas*) derived with notched noise," J. Acoust. Soc. Am. **112**, 322–328.

Fish, J. F., Johnson, C. S., and Ljungblad, D. K. (**1976**). "Sonar target discrimination by instrumental human divers," J. Acoust. Soc. Am. **59**, 602–606.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Helweg, D. A., Roitblat, H. L., Nachtigall, P. E., and Hautus, M. J. (**1996**). "Recognition of aspect-dependent three-dimensional objects by an echolocating Atlantic bottlenose dolphin," J. Exp. Psychol. **22**, 19–31.

Herman, L. M., and Arbeit, W. R. (**1972**). "Frequency difference limens in the bottlenose dolphin: 1-70 KC/S," J. Aud Res. **2**, 109–120.

Hudspeth, A. J. (**1983**). "The hair cells of the inner ear," Sci. Am. (Int. Ed.) **248**, 54–64.

Johnson, C. S. (**1967**). "Sound detection thresholds in marine mammals," in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, Oxford), pp. 247–260.

Johnson, C. S. (**1968**). "Masked tonal thresholds in the bottlenosed porpoise," J. Acoust. Soc. Am. **44**, 965–967.

Johnson, R. A., Moore, P. W. B., Stoermer, M. W., Pawloski, J. L., and Anderson, L. C. (**1988**). "Temporal order discrimination within the dolphin critical interval," in *Animal Sonar Processes and Performance*, edited by P. E. Nachtigall and P. W. B. Moore (Plenum, New York), pp. 317–321.

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.

Lemonds, D. W. (**1999**). "Auditory filter shapes in an Atlantic bottlenose dolphin (Tursiops truncatus)." Unpublished dissertation. University of Hawaii, Manoa.

Merzenich, M. M., and Schreiner, C. E. (**1992**). "Mammalian auditory cortex - Some comparative observations," in *The Evolutionary Biology of Hearing*, edited by Webster, D. B., Fay, R. A., Popper, A. N., Long, Glenis. R. (Springer, New York), pp. 673–689.

Moore, P. W. B., Hall, R. W., Friedl, W. A., and Nachtigall, P. E. (**1984**). "The critical interval in dolphin echolocation: What is it?," J. Acoust. Soc. Am. **76**, 314–317.

Moore, P. W. B., Pawloski, D. A., and Dankiewicz, L. (**1995**). "Interaural time and intensity difference thresholds in the bottlenose dolphin (*Tursiops truncatus*)," in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden, The Netherlands).

Nachtigall, P. E. (**1980**). "Odontocete echolocation performance on object size, shape and material," in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 71–95.

Pack, A. A., Herman, L. M., Hoffmann-Kuhnt, M., and Branstetter, B. K. (**2002**). "The object behind the echo: Dolphins (*Tursiops truncatus*) perceive object shape globally through echolocation," Behav. Processes **58**, 1–26.

Patterson, R. D. (**1994**). "The sound of a sinusoid: Spectral models," J. Acoust. Soc. Am., **96**, 1409–1418.

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. H. (**1992**). *Complex Sounds and Auditory Images* (Pergamon, Oxford).

Regan, M. P. (**1994**). "Linear half-wave rectification of modulated sinusoids," Appl. Math. Comput. **62**, 61–79.

Renaud, D. L., and Popper, A. N. (**1975**). "Sound localization by the bottlenose porpoise *Tursiops truncatus*," J. Exp. Biol. **63**, 569–585.

Roitblat, H. L., Ketten, D., Au, W. W. L., and Nachtigall, P. E. (**1996**). "A computational model of early stages of dolphin hearing," J. Acoust. Soc. Am. **100**, p. 2643.

Roitblat, H. L., Moore, P. W. B., Helweg, D. A., and Nachtigall, P. E. (**1993**). "Representation of acoustic information in a biomimetic neural network," in *From Animals to Animats 2: Simulation of Adaptive Behavior*, edited by J. A. Meyer, S. W. Wilson, and H. L. Roitblat (MIT, Cambridge, MA), pp. 90–99.

Slaney, M. (**1993**). "An efficient implementation of the Patterson-Holdsworth filter bank," Apple Technical Report No. 35, Advanced Technology Group, Apple Computer, Inc., Cupertino, CA.

Southworth, C., and Berg, B. G. (**1995**). "Multiple cues for the discrimination of narrow-band sounds," J. Acoust. Soc. Am. **98**, 2486–2492.

Vel'min, V. A., and Dubrovskii, N. A. (**1976**). "The critical interval of active hearing in dolphins," Sov. Phys. Acoust. **2**, 351–352.

Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based on modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.

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, New York), pp. 116–154.

Zorikov, T. V., and Dubrovski, N. A. (**2003**). "Echo-processing in bottlenose dolphins," Oceans 2003 Conference, San Diego, CA, pp. 320–326.

Attenuation in trabecular bone: A comparison between numerical simulation and experimental results in human femur

Emmanuel Bossy
Laboratoire Photons et Matière, ESPCI, CNRS UPR 5, F-75005 France and Université Pierre et Marie Curie-Paris 6, F-75005 France

Pascal Laugier
CNRS, UMR7623 LIP, Paris, F-75006 France and Université Pierre et Marie Curie-Paris 6, UMR7623, Paris, F-75005 France

Françoise Peyrin
CREATIS, UMR CNRS 5515, INSERM U630, 69621 Villeurbanne, Cedex, France and ESRF, BP 220, 38043 Grenoble, Cedex, France

Frédéric Padilla[a)]
CNRS, UMR7623 LIP, Paris, F-75006 France and Université Pierre et Marie Curie-Paris 6, UMR7623, Paris, F-75005 France

Numerical simulations (finite-difference time domain) are compared to experimental results of ultrasound wave propagation through human trabecular bones. Three-dimensional high-resolution microcomputed tomography reconstructions served as input geometry for the simulation. The numerical simulation took into account scattering, but not absorption. Simulated and experimental values of the attenuation coefficients (α, dB/cm) and the normalized broadband ultrasound attenuation (nBUA, dB/cm/MHz) were measured and compared on a set of 28 samples. While experimental and simulated nBUA values were highly correlated ($R^2 = 0.83$), and showed a similar dependence with bone volume fraction, the simulation correctly predicted experimental nBUA values only for low bone volume fraction (BV/TV). Attenuation coefficients were underestimated by the simulation. The absolute difference between experimental and simulated α values increased with both BV/TV and frequency. As a function of frequency, the relative difference between experimental and simulated α values decreased from 60% around 400 kHz to 30% around 1.2 MHz. Under the assumption that the observed discrepancy expresses the effect of the absorption, our results suggests that nBUA and its dependence on BV/TV can be mostly explained by scattering, and that the relative contribution of scattering to α increases with frequency, becoming predominant (>50 %) over absorption for frequencies above 600 kHz. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2766779]

PACS number(s): 43.80.Qf, 43.20.Fn, 43.80.Jz, 43.20.Hq [FD]   Pages: 2469–2475

I. INTRODUCTION

Quantitative ultrasound (QUS) has proven to be widely and clinically useful to assess bone and predict fracture risk [Gluer *et al.*, 2004; Khaw *et al.*, 2004; Marin *et al.*, 2006; Stewart *et al.*, 2006]. The commonest used QUS modality is applied to the measurement of cancellous bone at the calcaneus, although recent technology developments aim at also measuring the proximal femur [Barkmann *et al.*, 2007].

QUS in principle should allow an advanced assessment of bone strength since the ultrasound propagation characteristics are principally determined by structural and material properties of the propagation medium. However, this potential advantage of ultrasound has not been fully exploited so far because the underlying physical theory is still not completely understood despite recent advances in theoretical models which better explain the relationship between ultrasound parameters and structural and material bone properties. Although our knowledge has progressively improved, with recent studies investigating the validity of Biot's models [Lee *et al.*, 2003; Fellah *et al.*, 2004; Hosokawa, 2006], Schoenberg's theory [Hughes *et al.*, 1999] and the role of scattering applied to cancellous bone [Wear, 1999; Padilla *et al.*, 2003], there is still no unique comprehensive theoretical framework to account for the overall experimental observations.

The acoustic modeling of cancellous bone is complex because cancellous bone is a highly porous material consisting of a solid framework of mineralized collagen of connected rod- or plate-like elements and soft tissue, i.e., bone marrow filling the pore spaces. The trabecular alignment is determined by the stresses applied to the bone and varies with location. In younger ages the porosity is about 70% and

---

a)Author to whom correspondence should be addressed. Electronic mail: padilla@lip.bhdc.jussieu.fr

the elements are arranged in a plate-like network with a few rods. With increasing age more and more plates are perforated, porosity increases, and a more rod-like network is formed. The average size of the trabecular elements varies between 50 and 150 μm, the average separation between 0.5 and 2 mm [Chaffai *et al.*, 2002]. These distances are close to the ultrasonic wavelength (1.5 mm at 1 MHz).

Because of the technical difficulty to analytically model the extremely complex ultrasonic field resulting from the interaction of an incident wave with cancellous bone while taking into account the full complexity of the bone structure, researchers have more recently turned to bone computational models based on finite-difference simulations. These bone finite-difference studies, using either a generic model of the cancellous structure [Hosokawa, 2005, 2006] – which is limited by the assumptions regarding the morphology of the structure—or imaging the structure with microcomputed tomography [Luo *et al.*, 1999; Bossy *et al.*, 2005; Haiat *et al.*, 2007], provided valuable insights into the relationships between ultrasound propagation characteristics and bone properties. Two-dimensional (2D) finite-difference time domain (FDTD) simulation studies have provided information on how density and architecture affect ultrasound propagation [Luo *et al.*, 1999; Hosokawa, 2005, 2006]. Coupling numerical simulation with real three-dimensional (3D) microstructural models of trabecular bone reconstructed from synchrotron radiation microtomography (SR-μCT) has been shown to be a powerful tool to investigate the physics of ultrasound propagation in trabecular structures [Bossy *et al.*, 2005] and to quantitatively assess the effects of bone volume fraction, microstructure and material properties on ultrasonic attenuation and velocity [Haiat *et al.*, 2007]. Simulated results such as the findings that attenuation varies linearly with frequency, slope of frequency-dependent attenuation and sound velocity increase quasi linearly with the bone volume fraction, majority of samples shows a negative velocity dispersion and two compression waves may propagate upon certain trabecular orientations agreed well with experimental measurements and impressively demonstrated the power of these computational tools [Bossy *et al.*, 2005].

However, these numerical studies involve model assumptions regarding the material properties that are required to model bone. To assess the ability of such simulation to predict the outcome of a real experimental measurement, a face-to-face comparison between simulated and experimental measurements is needed. The goal of the present study was to provide such a face-to-face comparison, between 3D FDTD-based simulations and experiments performed on the same set of samples. Toward this end, 3D high-resolution μCT-based finite-difference bone models were developed for a cohort of human femoral specimens. Accurate modeling of 3D cancellous bone microstructure using μCT and direct comparison of simulated and experimental data distinguishes this study from previous ones. Our objectives were to compare the predicted and measured frequency-dependent attenuation through trabecular bone, to provide insight into the model assumptions on the outcomes of the numerical predictions and to gain a better understanding of the physical origin of the attenuation in trabecular bone.
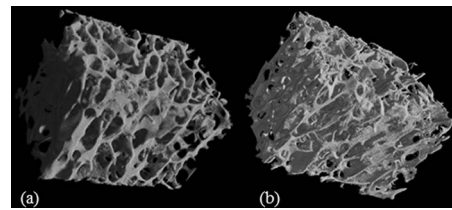

FIG. 1. Three-dimensional view of synchrotron microtomographic reconstruction of a typical dense (a) and porous (b) trabecular samples.

II. MATERIAL AND METHODS

A. Specimen

Twenty-eight human right femur specimens were dissected from fresh cadavers (nine women, mean age of 78 ± 11 years old, age range 55–91; and 21 men, mean age of 71 ± 10 years old, range 45–86). Ethical approval for collection of samples was granted by the Human Ethics Committee of the Institute of Anatomy at the University René Descartes (Paris, France). The tissue donors or their legal guardians provided informed written consent to provide their tissues for investigation, in accord with legal clauses stated in the French Code of Public Health (Code de la Santé Publique Français).

None of the donors underwent femoral prosthesis surgery. The skeletal status of the donors was undocumented. The specimens were stored at $-20\,°$C after dissection. The bone specimens were prepared in several steps. The femoral head and the diaphysis were first removed leaving the proximal part of the femur (neck and trochanter). Slices of trabecular bone with parallel faces and 10 mm thickness were cut from the proximal femur, in the plane defined by the cephalo-caudal and medio-lateral directions using a circular saw under continuous irrigation. The samples thickness was measured using an electronic caliper. The specimens were defatted using a dichloromethane ($C_2H_2Cl_2$) solution (immersion for two or three days) and water jet. To remove any air bubbles trapped in the defatted samples, they were vacuum-degassed underwater in a desiccators before ultrasonic measurements.

B. Computational bone models

The numerical computations were performed on 3D numerical models of trabecular structures [Bossy *et al.*, 2005]. Toward this end, cylindrical cores were machined in the prepared specimens (typically 10 mm in height and 8 mm in diameter). The 3D microarchitecture of parallelepipedic domains with dimensions 5.6×5.6 mm^2 in the transverse dimensions (X,Y) and 10 mm along the Z axis was reconstructed for each specimen measured using synchrotron radiation microcomputed tomography (SR-μCT) [Salome *et al.*, 1999] at the European Synchrotron Radiation Facility, France, with a voxel size of 10 μm in the three directions of space (Fig. 1). From the initial resolution of 10 μm, the volumes were downsampled to a voxel size of 30 μm to match memory requirements on a standard desktop PC while maintaining a reasonably accurate representation of the trabecular structure. The bone volume fraction, defined as the total bone volume normalized to the total specimen volume (*BV/TV*),

2470   J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Bossy *et al.*: Attenuation in trabecular bone

was obtained for each cylindrical core from reconstructed SR-μCT 3D datasets.

Simulations of wave propagation through trabecular bone were performed using SimSonic, a software developed by our group based on a finite-difference time domain (FDTD) algorithm [Bossy et al., 2004]. A numerical solution to the 3D linear elastic wave propagation equation is computed in nonattenuating fluid and solid materials. Perfectly matched layers [Collino and Tsogka, 2001] have been implemented, avoiding unphysical reflections from the boundaries of the simulation mesh. The geometry of the propagation media is defined by a simple 3D regular mapping of the mechanical properties of the two media (water and bone). Complex phenomena such as reflection, refraction, mode conversion at the numerous bone interfaces and therefore multiple scattering in cancellous bone, are fully taken into account by the simulation tool. However, the version of the software that was used does not account for absorption phenomena such as viscous effects.

To simulate the propagation of plane waves in the Z direction through samples of limited transverse dimensions, symmetric boundary conditions were applied on the ZX and YZ faces of the simulation domain. The ultrasonic pressure source was defined over the whole XY plane at $z=0$ mm. Therefore, the simulation modeled the propagation of plane waves through trabecular structures periodized by the boundary conditions in the X and Y directions. A 5.6×5.6 mm^2 plane receiver was set at $z=10$ mm, recording the pressure averaged over its whole surface. The source emits a broadband Gaussian pressure pulse centered at 1 MHz, close to the measured reference signal transmitted in water.

For all computations, the interstitial phase was considered to be nonviscous water, with density of 1 g/cm^3 and acoustic wave velocity of 1500 m/s. Bone tissue was assumed to be isotropic and nonabsorbing, with a constant density of 1.85 g/cm^3 and longitudinal wave velocity of 4000 m/s and transverse wave velocity of 1800 m/s, which are the typical elastic characteristics of cortical bone found in the literature [Njeh et al., 1999].

Both the spectra of the signal transmitted through bone and of a reference signal transmitted through water were required to compute the frequency-dependent attenuation coefficient and the speed of sound. Assuming an incident plane wave, the spectrum of the transmitted signal may be expressed as a function of the spectrum of the reference signal as

$$S_{\text{bone}}(f) = S_{\text{ref}}(f) \times T^2 \times e^{-\alpha(f)d} \times e^{-i(k_0-k)d}, \quad (1)$$

where f is the frequency, T is the overall transmission coefficient accounting for the transmission losses at both the water to sample interface and sample to water interface, $e^{-\alpha(f)d}$ is the transfer function for attenuation, $\alpha(f)$ being the frequency-dependent attenuation coefficient of the bone specimen, k_0 and k are the effective wave numbers in water and cancellous bone, respectively, and d is the thickness of the specimens. Following the same assumptions used to derive experimental values, T was assumed to be close to one and the frequency-dependent attenuation coefficient of the bone specimen expressed in cm^{-1} was derived from the log

ratio of the amplitude spectra [Laugier et al., 1997]

$$\alpha(f) = -\frac{1}{d} \ln \left| \frac{S_{\text{bone}}(f)}{S_{\text{ref}}(f)} \right|. \quad (2)$$

In the following, values for attenuation coefficient are given in dB/cm (rather than cm^{-1}) as the unit currently used in the bone characterization field. As reported previously, both the experimental [Chaffai et al., 2000] and simulated [Bossy et al., 2005] values of the attenuation coefficient were found to be linear over a wide frequency bandwidth ranging from 0.4 to 1.2 MHz, allowing calculation of the slope of the attenuation coefficient as a function of frequency. The $nBUA$ (normalized broadband ultrasound attenuation), representing this slope, was derived as the slope of a linear fit of the attenuation coefficient as a function of frequency in the 0.4–1.2 MHz frequency bandwidth.

C. Ultrasound measurements

Prior to synchrotron investigation, ultrasound transmission experiments were performed for each sample. The experimental protocol has been described elsewhere [Chaffai et al., 2000]. Briefly, ultrasonic measurements were performed in immersion using a pair of focused broadband transducers (1 MHz center frequency, 29 mm in diameter, focal length of 35 mm, V391, Panametrics Inc, Waltham, MA). The −6 dB beam width was approximately 3 mm. The −20 dB frequency bandwidth was 0.4–1.2 MHz. Emission and reception of the signals were performed using a plug-in ultrasonic pulser receiver, amplifier and digitizer (SFT 4001H PCI, Sofratest, Ecquevilly, France). Both transducers were mounted coaxially, separated by twice the focal length, in a through-transmission normal incidence configuration. The measured specimen was placed in the focal zone. Two stepping motors mounted on a crossed slide assembly moved the specimen transversally in the ultrasonic beam. Transmitted radio-frequency (rf) signals were recorded along 2D scans in steps of 1 mm (Motion Controller – MM4006 – Newport, Irvine, CA). The size of the scans were chosen to fit the sample size. Radio-frequency signals were amplified, time averaged and digitized at 60 MHz. Ultrasound measurements were made at room temperature at approximately 17 ° C. The attenuation coefficient as a function of frequency was estimated from the experimental signals in the same manner as from the simulated signals.

Image analysis was made using IMAGE J v1.29 free software. This software was used to select site-matched identical regions of interest (ROIs) on QUS 2D scans, matching the location of cylindrical cores extraction. Circular site-matched ROIs, 7 mm in diameter, were selected from QUS images in order to fit at best the cylindrical cores studied using SR-μCT. The average value of $nBUA$ was then calculated for each selected ROI.

III. RESULTS

Figure 2 shows the reference signals through water used for the experimental and simulated measurements, and two typical experimental and simulated signals both recorded on the same specimen. For this specimen (No. 296), the $nBUA$
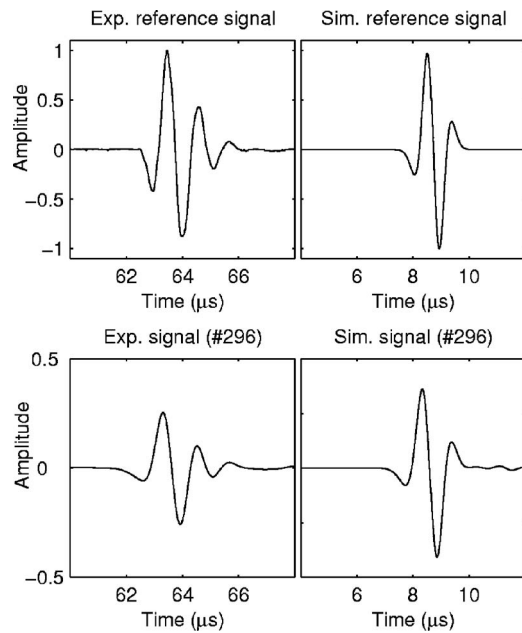
FIG. 2. Reference signal in water (top) from experiment (left) and simulation (right). Signal after propagation through specimen number No. 296 (bottom) from experiment (left) and simulation (right).

values were found to be 12.4 and 11.1 dB/cm/MHz for the experiment and simulation respectively. It can be seen that the transmitted signals have been low-pass filtered during their propagation. As already reported in previous publications [Bossy *et al.*, 2005; Jenson *et al.*, 2006], the attenuation in experiments and simulations varies linearly with frequency for all the specimens. For both experimental and simulated data, the determination coefficient (R^2) obtained by linearly fitting the dependence of the attenuation as a function of frequency in the 0.4–1.2 MHz bandwidth was found to be greater than 0.98 for all specimens except for one (for which R^2 was found to be 0.99 and 0.91 on experimental and simulated data, respectively).

Figure 3 plots the experimental and simulated *nBUA* as a function of *BV/TV*. Both sets of *nBUA* values were found to be highly correlated with *BV/TV* ($R^2 = 0.84$ for experiments and $R^2 = 0.90$ for simulation, both $p < 10^{-3}$). The regression
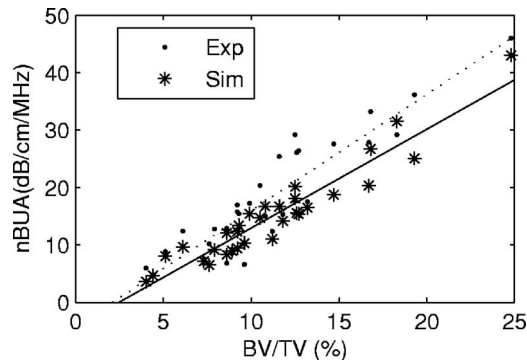

FIG. 3. Simulated and experimental nBUA in dB/cm/MHz as a function of *BV/TV*. The dashed line is the regression line for the experimental points ($nBUA_{experiment} = 2 * BV/TV(\%) - 4.2$ dB/cm/MHz, $R^2 = 0.84$, $p < 10^{-5}$, RMSE = 2.6 dB/cm/MHz). The plain line is the regression line for the simulated points ($nBUA_{simulation} = 1.7 * BV/TV(\%) - 4.3$ dB/cm/MHz, $R^2 = 0.90$, $p < 10^{-5}$ RMSE = 4.1 dB/cm/MHz).
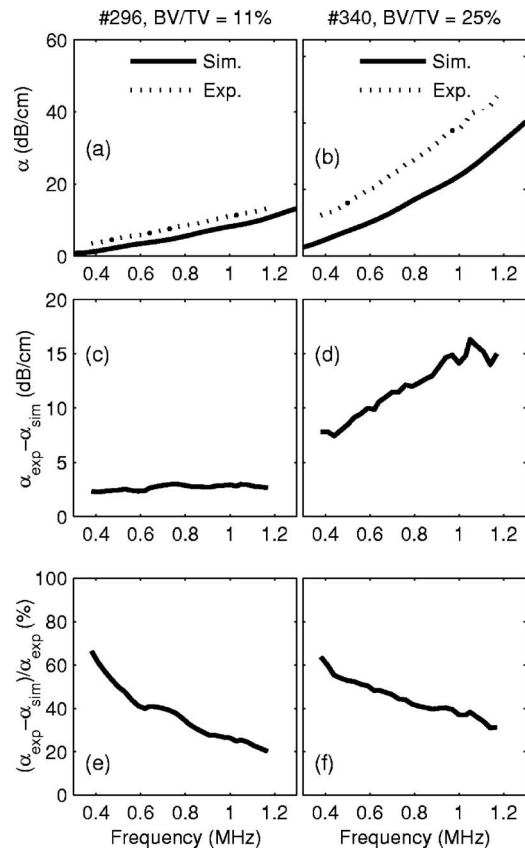

FIG. 4. Attenuation coefficients in dB/cm as a function of frequency for a specimen with low *BV/TV* (No. 296, *BV/TV* = 11%, left) and a specimen with high *BV/TV* (No. 340, *BV/TV* = 25%, right). (a) and (b): experimental and simulated attenuation coefficients. (c) and (d): absolute difference between experimental and simulated attenuation coefficients. (e) and (f): relative difference between experimental and simulated attenuation coefficients.

equations shown in Fig. 3 indicate a gradient of simulated *nBUA* versus *BV/TV* of 1.7 dB/cm/MHz per % of *BV/TV*, which is similar in magnitude to what was found with experimental data (2.0 dB/cm/MHz per % of *BV/TV*). An overall qualitative agreement was found between simulated and experimental *nBUA* (a linear regression yielded $nBUA_{experiment} = 1.12$. $nBUA_{simulation} + 1.6$ dB/cm/MHz, $R^2 = 0.83$, $p < 10^{-5}$, RMSE (root mean squared error) = 4.1 dB/cm/MHz). Nevertheless, Fig. 3 indicates that the simulation correctly predicts *nBUA* values only for specimens with relatively low *BV/TV*, and there is, as a rule of thumb, a slight trend to underestimate *nBUA* values as *BV/TV* is increasing. This is also indicated by the gradient of *nBUA* which is steeper in the experimental observations compared to the simulations.

To better understand the differences between simulated and experimental values of *nBUA*, the comparison was pushed further by analyzing the difference at each frequency between the experimental and simulated attenuation coefficients. Figure 4 plots the frequency-dependent attenuation coefficients for two typical specimens, characterized respectively by low and a high *BV/TV* values. For specimen No. 296, the *BV/TV* value is 11% and the experimental and simulated values of *nBUA* were found to be 11 and 12 dB/(cm MHz), respectively. For the specimen No. 340, the *BV/TV* value is 25% and the experimental and simulated
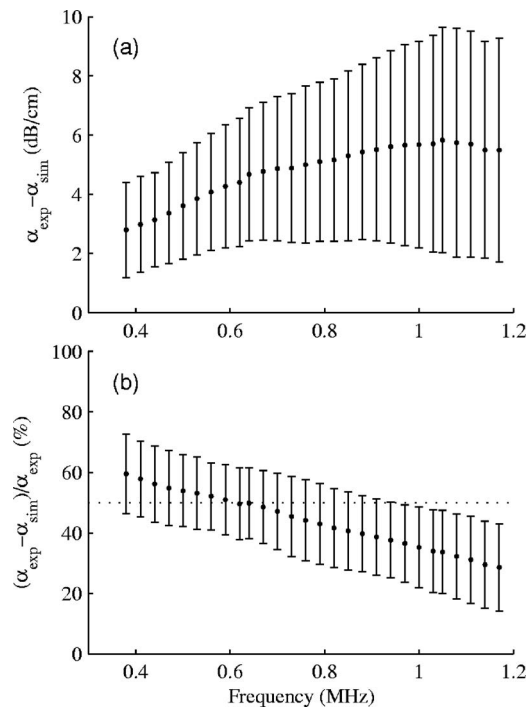
FIG. 5. Average value of the absolute (a) and relative (b) difference between experimental and simulated attenuation coefficients as a function of the frequency. The error bars denote the standard deviation over the group of specimens.

values of *nBUA* were found to be 43 and 46 dB/(cm MHz), respectively. The top plots illustrate the point already noted in Fig. 3: while the simulation provides a close prediction of *nBUA* for specimens at low *BV/TV* value, it underestimates *nBUA* for the specimens with high *BV/TV* values. Moreover, Fig. 4 shows that this trend is not only true for the *nBUA* values, but also for the frequency-dependent attenuation coefficient. For specimen No. 340, with high *BV/TV* value, the absolute difference between simulated and experimental attenuation coefficients originates both from a constant offset (ordinate at the origin) and from the difference of slope, as shown by the middle plot. In this case, the higher the frequency, the larger the absolute difference between experimental and simulated attenuation data. However, it is interesting to notice that the trend is inverted for the relative difference between experimental and simulated attenuation coefficients: the relative difference is a decreasing function of frequency (Figs. 4(e) and 4(f)). This behavior has been confirmed on average for all the specimens, as shown in Fig. 5: plot (a) shows that the average absolute difference between experimental and simulated attenuation coefficients is positive and an increasing function of frequency. Plot (b) shows that the average relative difference between experimental and simulated attenuation coefficients is a decreasing function of frequency.

IV. DISCUSSION

The results above show that, although the simulated attenuation coefficients and *nBUA* values are in qualitative agreements with the corresponding experimental values in terms of their dependency with frequency and *BV/TV*, the

simulations underestimate the attenuation coefficient as well as its slope as a function of frequency, especially for specimens with high *BV/TV*. Several reasons may explain this discrepancy.

First, the computational model relies on several assumptions which must be considered as potential sources of random and systematic discrepancies between simulated and experimental results. The simulations are free of noise, measurement errors or other factors (anisotropy and heterogeneity of material properties) that are likely to play a role in the variability of experimental results. For example, the true local material properties were not known (generic values are used instead) and the bone tissue was considered as mechanically isotropic. We have demonstrated in an complement simulation study [Haiat *et al.*, 2007] that *nBUA* could vary at most by 10% in response to a 20%–30% variation in material properties (longitudinal and shear waves velocities, material density). Fluctuations in material properties between or within the specimens could therefore have a non-negligible effect on *nBUA*. Nevertheless, this effect is unlikely to explain the difference between simulated and experimental values that can reach 30%. Effects such as heterogeneity and anisotropy of the material properties on the amplitude of the attenuation coefficient remain to be quantified.

Second, and most importantly, the FDTD simulation does not include any absorption mechanisms. The measured attenuation in virtual samples is caused by scattering only, whereas absorption in real cancellous bone can contribute to the total attenuation through several mechanisms. These include absorption in the solid framework (bone tissue) and in the fluid, and losses by viscous friction at the bone/fluid interface. Absorption is present in cortical bone tissue, and probably finds its origin in the viscoelastic nature of the tissue [Sasaki *et al.*, 1993]. For longitudinal waves, attenuation of several dB/cm was measured at frequencies sufficiently low to avoid scattering by structural elements [Lakes *et al.*, 1986; Wu and Cubberley, 1997]. For the shear waves, the attenuation was found to be even higher (approximately two times higher) than for longitudinal waves [Lakes *et al.*, 1986; Wu and Cubberley, 1997]. Although values for absorption coefficients in cancellous bone tissue have not been reported yet to date, similarities in composition and structure of trabecular and cortical bone tissues suggest that absorption losses of longitudinal and shear waves may be close in cortical and trabecular bone tissues. Therefore, both longitudinal and shear waves, which are excited by mode conversions at the water/bone interfaces, are likely to be absorbed experimentally during their propagation in the solid tissue and these absorption mechanisms, which are not accounted for in the FDTD simulation, are likely to contribute to the overall experimental attenuation losses. Viscous friction at the bone/fluid interfaces as described in the context of Biot theory [Fellah *et al.*, 2004], and viscous absorption in the fluid are other avenues to consider to explain the discrepancy between simulated and measured attenuation.

Although neglecting the absorption is a limitation of the simulation study, our results provide insights into the respective contribution of absorption and scattering to the total attenuation. Assuming that scattering is properly accounted for

in the simulation and that the absence of absorption losses in the simulations is the main source of discrepancy between simulation and experiments, then the absolute difference between experimental and simulated attenuation coefficients represented in Figs. 4 and 5 can be interpreted as being simply the experimental absorption coefficient. First, results from Figs. 3, 4(a), and 4(b) suggest that the contribution to total losses of absorption losses increases with *BV/TV*, as both the difference between experimental and simulated attenuation coefficients and *nBUA* values increase with *BV/TV*. Second, considering the increase in the *absolute* difference between experimental and simulated attenuation coefficients when the frequency increases (Figs. 4(c), 4(d), and 5(a)), our results suggest that absorption losses also increase with frequency. Third, the fact that the *relative* difference between experimental and simulated attenuation coefficients decreases with frequency (Figs. 4(e), 4(f), and 5(b)) suggests that scattering losses increase faster with frequency than absorption losses do. Indeed, Fig. 5(b) indicates that in the MHz frequency range, scattering can predict approximately half of the total attenuation coefficient. More precisely, at 400 kHz, in the low frequency region of the available bandwidth, scattering represents approximately 40% of the total attenuation and at 1.2 MHz, in the high frequency region of the available bandwidth, scattering represents approximately 70% of the total attenuation.

The assumption that the absence of absorption losses in the simulations is responsible for the difference between simulation and measurements is supported by the fact that fluctuations in material properties cannot explain the discrepancy between simulated and experimental attenuation coefficient, as discussed previously. Further insights gained from future simulations including absorption should contribute to better interpretation of *nBUA* variations with bone volume fraction. It would be valuable to incorporate anisotropy and inhomogeneity of material properties in future studies to enhance final conclusions.

V. CONCLUSIONS

To summarize, our results demonstrate that numerical simulation can predict nBUA values close to the experimental ones especially for specimens with low bone volume fraction, and can predict accurately the variations of nBUA as a function of bone volume fraction.

Several important points were highlighted in this study: (1) scattering is probably the main contribution to *nBUA* for bone specimens with low *BV/TV*, but scattering alone is not sufficient to explain total attenuation in denser specimens in which cases other attenuation mechanisms such as absorption may be significant. (2) a significant part of the attenuation is explained by scattering. (3) the total attenuation is underestimated in the simulation. The absolute bias tends to increase as a function of both frequency and bone volume fraction as a result of increasing absorption. (4) Under the assumption that the discrepancy observed between simulated and experimental values reflects the effect of the absorption, not taken into account by the simulation, our results also suggest that the relative contribution of scattering to the total

attenuation losses tends to increase with frequency as a result of a steeper increase of scattering with frequency compared to absorption mechanisms.

Barkmann, R., Laugier, P., Moser, U., Dencks, S., Padilla, F., Haiat, G., Heller, M., and Gluer, C. C. (**2007**). "A method for the estimation of femoral bone mineral density from variables of ultrasound transmission through the human femur," Bone (N.Y.) **40**, 37–44.

Bossy, E., Padilla, F., Peyrin, F., and Laugier, P. (**2005**). "Three-dimensional simulation of ultrasound propagation through trabecular bone structures measured by synchrotron microtomography," Phys. Med. Biol. **50**, 5545–5556.

Bossy, E., Talmant, M., and Laugier, P. (**2004**). "Three-dimensional simulations of ultrasonic axial transmission velocity measurement on cortical bone models," J. Acoust. Soc. Am. **115**, 2314–2324.

Chaffai, S., Padilla, F., Berger, G., and Laugier, P. (**2000**). "In vitro measurement of the frequency-dependent attenuation in cancellous bone between 0.2 and 2 MHz," J. Acoust. Soc. Am. **108**, 1281–1289.

Chaffai, S., Peyrin, F., Nuzzo, S., Porcher, R., Berger, G., and Laugier, P. (**2002**). "Ultrasonic characterization of human cancellous bone using transmission and backscatter measurements: Relationships to density and microstructure," Bone (N.Y.) **30**, 229–237.

Collino, F., and Tsogka, C. (**2001**). "Application of the PML absorbing layer model to the linear elastodynamic problem in anisotropic heterogeneous media," Geophysics **66**, 294–307.

Fellah, Z. E. A., Chapelon, J. Y., Berger, S., Lauriks, W., and Depollier, C. (**2004**). "Ultrasonic wave propagation in human cancellous bone: Application of Biot theory," J. Acoust. Soc. Am. **116**, 61–73.

Gluer, C. C., Eastell, R., Reid, D. M., Felsenberg, D., Roux, C., Barkmann, R., Timm, W., Blenk, T., Armbrecht, G., Stewart, A., Clowes, J., Thomasius, F. E., and Kolta, S. (**2004**). "Association of five quantitative ultrasound devices and bone densitometry with osteoporotic vertebral fractures in a population-based sample: The OPUS study," J. Bone Miner. Res. **19**, 782–793.

Haiat, G., Padilla, F., Peyrin, F., and Laugier, P. (**2007**). "Variation of ultrasonic parameters with microstructure and material properties of trabecular bone: A 3D model simulation," J. Bone Miner. Res. **22**, 665–674.

Hosokawa, A. (**2005**). "Simulation of ultrasound propagation through bovine cancellous bone using elastic and Biot's finite-difference time-domain methods," J. Acoust. Soc. Am. **118**, 1782–1789.

Hosokawa, A. (**2006**). "Ultrasonic pulse waves in cancellous bone analyzed by finite-difference time-domain methods," *Ultrasonics Proceedings of Ultrasonics International (UI'05) and World Congress on Ultrasonics (WCU)*, **44**, e227-e231.

Hughes, E. R., Leighton, T. G., Petley, G. W., and White, P. R. (**1999**). "Ultrasonic propagation in cancellous bone: A new stratified model," Ultrasound Med. Biol. **25**, 811–821.

Jenson, F., Padilla, F., Bousson, V., Bergot, C., Laredo, J. D., and Laugier, P. (**2006**). "In vitro ultrasonic characterization of human cancellous femoral bone using transmission and backscatter measurements: Relationships to bone mineral density," J. Acoust. Soc. Am. **119**, 654–663.

Khaw, K. T., Reeve, J., Luben, R., Bingham, S., Welch, A., Wareham, N., Oakes, S., and Day, N. (**2004**). "Prediction of total and hip fracture risk in men and women by quantitative ultrasound of the calcaneus: EPIC-Norfolk prospective population study," Lancet **363**, 197–202.

Lakes, R., Yoon, H. S., and Katz, J. L. (**1986**). "Ultrasonic wave propagation and attenuation in wet bone," J. Biomed. Eng. **8**, 143–148.

Laugier, P., Droin, P., LavalJeantet, A. M., and Berger, G. (**1997**). "In vitro assessment of the relationship between acoustic properties and bone mass density of the calcaneus by comparison of ultrasound parametric imaging and quantitative computed tomography," Bone (N.Y.) **20**, 157–165.

Lee, K. I., Roh, H. S., and Yoon, S. W. (**2003**). "Acoustic wave propagation in bovine cancellous bone: Application of the modified Biot-Attenborough model," J. Acoust. Soc. Am. **114**, 2284–2293.

Luo, G. M., Kaufman, J. J., Chiabrera, A., Bianco, B., Kinney, J. H., Haupt, D., Ryaby, J. T., and Siffert, R. S. (**1999**). "Computational methods for ultrasonic bone assessment," Ultrasound Med. Biol. **25**, 823–830.

Marin, F., Gonzalez-Macias, J., Diez-Perez, A., Palma, S., and Delgado-Rodriguez, M. (**2006**). "Relationship between bone quantitative ultrasound and fractures: A meta-analysis," J. Bone Miner. Res. **21**, 1126–1135.

Njeh, C., Hans, D., Fuerst, T., Glüer, C. C., and Genant, H. K. (**1999**). *Quantitative Ultrasound. Assessment of Osteoporosis and Bone Status* (Martin Dunitz, London).

Padilla, F., Peyrin, F., and Laugier, P. (**2003**). "Prediction of backscatter coefficient in trabecular bones using a numerical model of three-dimensional microstructure," J. Acoust. Soc. Am. **113**, 1122–1129.

Salome, M., Peyrin, F., Cloetens, P., Odet, C., Laval-Jeantet, A. M., Baruchel, J., and Spanne, P. (**1999**). "A synchrotron radiation microtomography system for the analysis of trabecular bone samples," Med. Phys. **26**, 2194–2204.

Sasaki, N., Nakayama, Y., Yoshikawa, M., and Enyo, A. (**1993**). "Stress-relaxation function of bone and bone-collagen," J. Biomech. **26**, 1369–1376.

Stewart, A., Kumar, V., and Reid, D. M. (**2006**). "Long-term fracture prediction by DXA and QUS: A 10-year prospective study," J. Bone Miner. Res. **21**, 413–418.

Wear, K. A. (**1999**). "Frequency dependence of ultrasonic backscatter from human trabecular bone: Theory and experiment," J. Acoust. Soc. Am. **106**, 3659–3664.

Wu, J. R., and Cubberley, F. (**1997**). "Measurement of velocity and attenuation of shear waves in bovine compact bone using ultrasonic spectroscopy," in Ultrasound Med. Biol. 129–134.

J. Acoust. Soc. Am., Vol. 122, No. 4, October 2007

Bossy *et al.*: Attenuation in trabecular bone   2475

Erratum: "5aMU4. Acoustics of music and voice in Jewish worship spaces" [J. Acoust. Soc. Am. 121(5), 3193–4 (2007)]

There is a publisher's error in the list of authors of this paper as it appeared in the Salt Lake City Meeting Program. The correct author list is as follows:

Mendel Kleiner (Chalmers Rm. Acoustics Group, Department of Appl. Acoust., Chalmers University of Technology, 41296 Gothenburg, Sweden, mendel.kleiner@chalmers.se) and David L. Klepper (Yeshivat Beit Orot, Mt. of Olives, Jerusalem 97400, Israel).