CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 126, No. 2 August 2009

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

Analysis of the part-pedaling effect in the piano

Heidi-Maria Lehtonen
Department of Signal Processing and Acoustics, Helsinki University of Technology, P.O. Box 3000, FI-02015 TKK, Espoo, Finland
heidi-maria.lehtonen@tkk.fi

Anders Askenfelt
Department of Speech, Music and Hearing, Royal Institute of Technology (KTH), Lindstedtsvägen 24, SE-10044 Stockholm, Sweden
andersa@speech.kth.se

Vesa Välimäki
Department of Signal Processing and Acoustics, Helsinki University of Technology, P.O. Box 3000, FI-02015 TKK, Espoo, Finland
vesa.valimaki@tkk.fi

Abstract: This letter reports basic acoustic phenomena related to part-pedaling in the piano. With part-pedaling, the piano tone can be divided into three distinct time intervals: initial free vibration, damper-string interaction, and final free vibration. Varying the distance of the damper from the string, the acoustic signal and the damper acceleration were measured for several piano tones. During the damper-string interaction, the piano tone decay is rapid and the timbre of the tone is affected by the nonlinear amplitude limitation of the string motion. During the final free decay, the string continues to vibrate freely with a lower decay rate.

PACS numbers: 43.75.Mn, 43.75.Zz [DKW]
Date Received: March 30, 2009 **Date Accepted:** June 4, 2009

1. Introduction

The sustain pedal is generally recognized as an essential part of the piano, and it is used to create different artistic expressions in piano performances.[1] Professional pianists apply different pedaling techniques, such as part-pedaling, vibration pedal, and pulsation pedal instead of just depressing and releasing the pedal.[1] Especially part-pedaling is used very often. Part-pedaling in this context means a common use of the sustain pedal where the pedal is not fully depressed, but pressed somewhere between the two extremes. In this letter, the effect of part-pedaling on piano tones is studied through recordings and signal analysis. Interesting questions are how the decay characteristics of the tones vary as a function of the depth of the sustain pedal and how the interaction between the damper and the string affects the sound. Earlier studies[2,3] concentrated on the differences in the two extreme positions: the sustain pedal is not used and it is fully depressed.

Physically informed sound synthesis of musical instruments has gained popularity during the past decades.[4] In order to produce an authentic piano sound with this technique, the underlying acoustical principles of the instrument need to be known. Although the acoustics of the piano are well documented in the literature (see, e.g., the introduction of Ref. 5), the effect of the pedals, and especially the use of different pedaling techniques, has been studied less. Algorithms for producing the full sustain-pedal effect in physics-based piano synthesizers have already been presented.[2,6,7] The fundamental goal of the present study was to obtain information to support the development of a sustain-pedal algorithm for a physics-based piano synthesizer.

Figure 1(b) illustrates how part-pedaling affects the sound waveform of the tone G3. Compared to the tone played without the sustain pedal [Fig. 1(a)], one notices that the decay includes a prolonged, soft tail after the key release, lasting for several seconds. This indicates

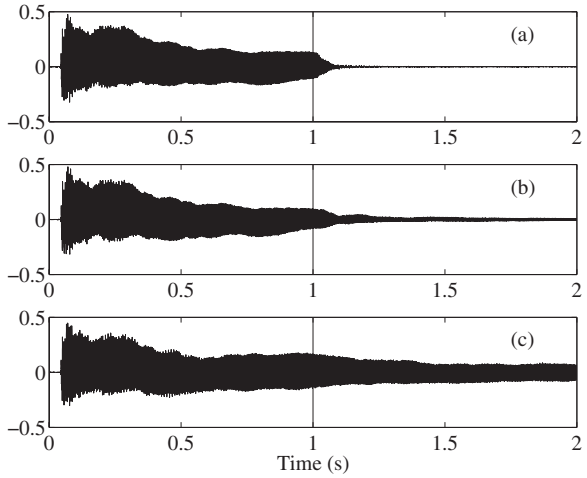


Fig. 1. Sound waveforms of the piano tone G3 (f_0=196.4 Hz) played (a) without the sustain pedal, (b) with part-pedaling, and (c) with full sustain pedal. The curves are aligned at key release (vertical lines), defined as the time instant when the key reached the upper position.

that the damper does not fully attenuate the vibrating string when the key is released. The vertical lines at 1.0 s show the time instant when the key has been fully released. In all three cases the key was released fast by suddenly removing the finger from the key. The example tones are available for listening at the companion web page of this article.[8]

2. Recordings and measurement setup

Six *mezzo forte* tones (C1, A1, G2, G3, D4, and A4) of a Steinway & Sons grand piano (model C, 224 cm) in concert condition were recorded in a rehearsal studio. The tones were selected so that all four damper shapes used in the instrument were represented: C1 was a one-string low bass tone with a $\wedge$ profile damper; A1 was a double-string bass tone with a $\vee$ profile damper; G2 and G3 were string triplets with double wedges; D4 was a string triplet with a double wedge on the half toward the agraffe and flat cushion toward the bridge; and A4 was a plain string triplet with a flat cushion. The tones were recorded with a Brüel & Kjær model 4003 omnidirectional condenser microphone (powered by a B&K 2812 power supply) positioned about 30 cm above the strings. In order to be able to synchronize the audio signal to the different phases of damper movement, a miniature accelerometer (B&K 4374, mass 0.7 g) was attached on top of the corresponding damper. The tones were played manually. The position of the dampers (height above the strings) was controlled through adjustment of the sustain-pedal rod nut instead of by pressing the sustain pedal with the foot.

Two types of recordings were carried out, both with accurate control of the damper position. First, only the dampers of the selected tones were regulated so that they did not touch the strings when the sustain pedal was not pressed down. This was done by adjusting the position of the damper wire in the damper block. The distance between the selected dampers and the corresponding string groups were then altered in steps by turning the adjustment nut on the sustain-pedal rod. In this way it was possible to control the damper position of the selected notes, while the rest of the strings remained fully damped. Starting from full damping, the distance between the dampers and the strings was increased in steps of 60° until the condition that was considered to correspond to a full sustain pedal was reached. This occurred after about two and three full turns in the treble and bass range, respectively. Each case was repeated three times. In the second recording session the damper regulation was returned to normal, i.e., the turning of the sustain-pedal rod nut made all dampers rise by the same amount. Compared to the previous case, this condition was closer to the normal use of the sustain pedal.

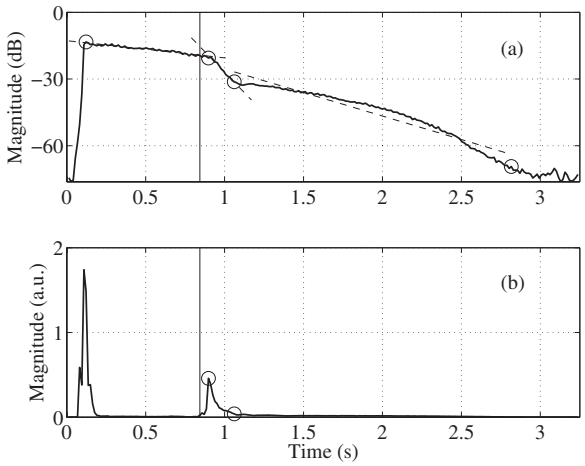


Fig. 2. (a) Log envelope of the microphone signal of the piano tone G3 using part-pedaling (solid line). The damper height adjustment was roughly 0.5 mm above full damping position. The dashed lines, indicating the decay times during initial free vibration, the damper-string interaction, and the final free vibration, were obtained by linear regression. The circles show the start and end points for the interval at which the lines have been fitted. The vertical line shows the approximate time instant when the key was released, determined from the shape of the magnitude of the damper acceleration signal. (b) The magnitude of the corresponding damper acceleration signal. The start and end points for the second time interval have been indicated with circles (see text).

The actual height of the damper was checked with a dial gauge, rigidly mounted on the metal plate inside the piano. In order to minimize the pressing force on the damper during measurements, a piece of copper foil was glued on top of the damper. Measurements were made by observing the electrical contact as the tip of the dial gauge touched the damper. It was found that one full turn corresponded to a 0.45 mm change in damper height. The adjustment step size of 60° corresponded to 0.08 ± 0.01 mm.

3. Analysis of decay characteristics

Decay characteristics were obtained from analysis of tones included in the first recording session, where all but the selected notes were fully damped in order to minimize the amplitude beating caused by sympathetic coupling to other strings. This choice leads to clearer results and reveals adequate information about how the interaction between the damper and the string group affects the decay process for different positions of the damper (corresponding to sustain-pedal depth in normal playing). Following the standard notation, decay time is defined here as the time that it takes for a tone to decay 60 dB.

When part-pedaling is used, the decay process can be divided into three time intervals. This is illustrated for tone G3 in Fig. 2, where panel (a) shows the log envelope of the microphone signal and panel (b) the corresponding envelope of the signal from the accelerometer that was attached to the top of the damper. The vertical line shows the time instant when the key is released, approximated from the shape of the magnitude of the damper acceleration signal. In this case, the damper height adjustment was approximately in the middle of the part-pedaling range, roughly corresponding to 0.5 mm above full damping position.

The first time interval, which is called the initial free vibration, is measured from the maximum of the sound signal log envelope to the second local maximum of the damper acceleration signal, which indicates when the damper stops its downward motion. The second time interval, during which an efficient damper-string interaction takes place, spans from the second local maximum of the damper acceleration signal to the point where the level of the acceleration signal has dropped 90% of the maximum. This level has been determined experimentally from the recorded data by comparing the shapes of the sound signal log envelopes and the magnitude of the damper acceleration signal. For the analyzed tones, this level corresponded well with the end of the second interval in the log envelope of the sound signal. The third time interval, the

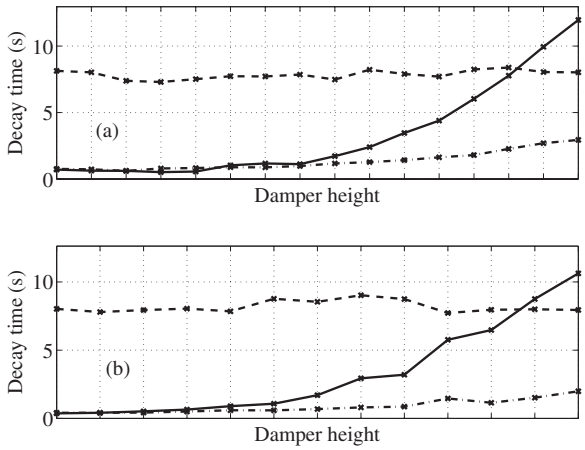


Fig. 3. Decay rates of the three time intervals of the decay (see text) as a function of the damper height adjustment for two piano tones (a) G2 (f_0=98.12 Hz) and (b) G3 (f_0=196.4 Hz): initial free vibration (dashed line), damper-string interaction (dash-dotted line), and final free vibration (solid line). Each adjustment step (grid line) corresponds to 0.08 mm.

final free vibration, is measured from the point where the previous interval ends and the free decay starts. This time interval ends when the log envelope of the microphone signal reaches a level which is 6 dB above the noise floor. The noise floor was measured using the nonlinear least-squares method proposed by Karjalainen *et al.*[9] The three phases of decay could be identified also in part-pedaling tones in the second set of recorded data where all dampers were lifted as the pedal was depressed. Only when the depth of the sustain pedal approaches full pedaling the clear structure of separate phases is lost, in agreement with observations in a previous work.[2]

The decay times for each of the three time intervals were measured using linear regression between the corresponding start and end points. Figure 3 shows the result of the decay time analysis for tones G2 and G3 as a function of damper height adjustment, each step (0.08 mm) indicated by the gridlines and data points. The decay times are averages across three repeated recordings. The three time intervals, the initial free vibration, the damper-string interaction, and the final free vibration are indicated with dashed, dash-dotted, and solid lines, respectively. The part-pedaling effect is seen to set in where the curves for the damper-string interaction and final free vibration start to separate.

As would be expected, the decay time of the initial free vibration remains uninfluenced by the damper adjustments as the damper is lifted to its upper position by the key during this interval. The decay time of the second interval, during which the damper and the string are in efficient contact and rapid damping takes place, increases slightly as a function of the distance between the damper and the string. This is also an expected result.

In addition, it can be seen that the decay time of the final free vibration increases as the distance between the damper and the string increases. A similar result was found also by Brauss.[1] This is expected, since the damper suppresses mainly the vertical polarization of vibration. After the interaction the string continues to vibrate mainly in the horizontal polarization, which is known to have a longer decay time than the vertical polarization.[10] The initial levels for the final free vibration increase 5–10 dB for G2 and G3, respectively, when the depth of the sustain pedal is increased from no pedal to full pedal. It is interesting to note that the typical part-pedaling effect occurs within a very limited range of damper height.

4. Analysis of the nonlinear effects of the damper-string interaction

When the key is released and part-pedaling is used, the damper hits the vibrating string causing a nonlinear limitation of the string amplitude. This is prominent especially in the bass range where the amplitude of the vibration is large. Nonlinear effects are known to transfer energy

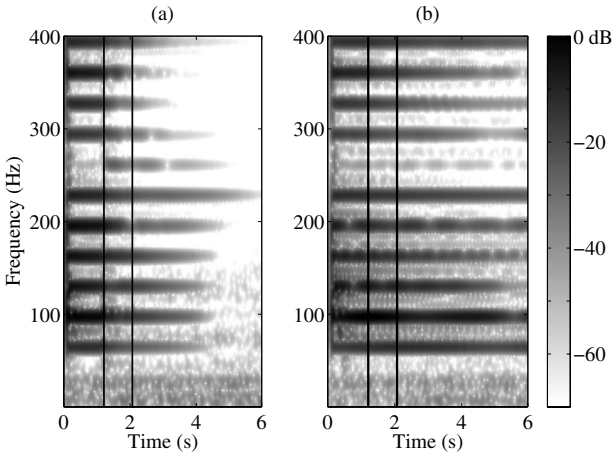


Fig. 4. Spectrograms of the tone C1 (f_0=32.54 Hz) played (a) with part-pedaling and (b) with full pedal. The vertical lines show the duration of the damper-string interaction when part-pedaling is used. The same interval is shown in (b) for reference.

from one mode to another in vibrating strings, for example, through tension modulation.[11] If the string has a nearly-missing mode, i.e., the string is excited close to a node of a string mode, this particular mode can gain energy from lower modes. In this sense, part-pedaling resembles the slapped bass effect,[12] where the frets or the fingerboard limit the amplitude of the string vibration.

This part of the analysis was performed on the second set of recorded tones with the dampers normally adjusted. The nonlinear effect caused by the damper interacting with the vibrating string was studied through spectrograms. Figures 4(a) and 4(b) show the result for the tone C1 when part-pedaling and full pedal, respectively, are used. The vertical lines indicate the duration of the damper-string interaction with part-pedaling determined from the damper acceleration signal, like in Fig. 2. The same interval is shown in Fig. 4(b) for reference. The spectrograms were computed using a Chebyshev window of length 300 ms with 250 ms overlap.

Figure 4(a) shows that the relations between the levels of the partials change during the decay when part-pedaling is used and influence the evolution of the timbre. The fundamental is very weak (not visible in the spectrogram), since it is not radiated efficiently by the soundboard. In addition, the hammer-string contact time is too short in the low bass to excite the lowest mode efficiently, and the striking position is unfavorable. Partials 2–7 are excited, but they decay fast compared to the case when the sustain pedal is fully depressed. Partial 8 is missing in Fig. 4(b) because the striking point is located at 192 mm from the agraffe, corresponding to almost exactly $L/8$, L=1573 mm being the string length. It can be seen in Fig. 4(a), however, that this missing mode gains energy when the damper-string interaction starts in the part-pedaling case. This is caused by the nonlinear limitation of the string amplitude. The tone C1 that is presented in Fig. 4 is available for listening at the companion web page of the article.[8]

For the midrange tones G2 and G3, the effect of energy transfer from lower to higher modes was not easy to observe, simply because there were no nearly-missing partials in the spectra. This is the case for many piano tones as the hammer striking position changes smoothly across the compass, and only for certain tones an efficient cancellation of some partials takes place. The timbre of the tone is still changing during the damper-string interaction, however, since some of the modes are attenuated more efficiently than others. This is explained by the damper position and length: if a particular mode has an antinode at the damper position it will die out quickly. On the other hand, if a mode has a node at the damper position and the length of the damper is short enough, the corresponding partial will keep ringing even if the adjacent partials are decaying fast. This effect can be seen in Fig. 4(a), where the seventh mode is decay-

ing slower compared to the lower modes. The damper covers the range 192–290 mm measured from the agraffe, and the seventh mode has a node at 1573/7=225 mm, which is close to the middle of the damper. Further, the length of the damper (98 mm) is less than a quarter of the wavelength of the partial 7 (112 mm). For the highest recorded tones D4 and A4, the described behavior was not observed since most of the partials above the partial 4 have an antinode at the damper position. In these tones the fundamental remains prominent after the damper-string interaction. For the examined mid- and high-range tones, it was also typical that the level of the tones decreased notably during the damper-string interaction.

5. Conclusions

This study presents new results on the effect of part-pedaling in the piano. When part-pedaling is used, the decay of piano tones can be divided into three distinct phases: the initial free vibration, the damper-string interaction, and the final free vibration. These three phases have different decay times. When the depth of the sustain pedal is increased from no pedal to full pedal, the decay time of the final free vibration is increased. In the bass range the nonlinear amplitude limitation causes energy transfer from the lower partials to higher partials, which can excite missing modes during the damper-string interaction.

Future work includes further investigation of the properties of the damper-string interaction. The behavior of the dampers affects the piano sound greatly, and especially for low tones the interaction with the string is a complicated process since the string vibrations cause vigorous vibrations in the damper as well.[13] Based on the present study, it can be concluded that a faithful sustain-pedal algorithm design must account for the effect of dampers in part-pedaling.

Acknowledgments

This work was enabled by the COST Action IC0601 on Sonic Interaction Design through a Short-Term Scientific Mission at Royal Institute of Technology, Stockholm, Sweden. The authors are grateful to piano technician Tore Persson for his expertise on damper regulation and to Balázs Bank for his comments related to this work. This research has been funded by the Academy of Finland (Project Nos. 122815 and 126310). The work of H.-M.L. is supported by the GETA Graduate School, Tekniikan edistämissäätiö, the Finnish Cultural Foundation, the Nokia Foundation, and Emil Aaltonen Foundation.

References and links

[1]H. Brauss, *The Singing Piano Tone or the Artistry of Pedaling* (Trafford, Victoria, Canada) (2006).
[2]H.-M. Lehtonen, H. Penttinen, J. Rauhala, and V. Välimäki, "Analysis and modeling of piano sustain-pedal effects," J. Acoust. Soc. Am. **122**, 1787–1797 (2007).
[3]A. Schutz, N. Bertin, D. T. M. Slock, B. David, and R. Badeau, "Piano "forte pedal" analysis and detection," in Proceedings of the 124th AES Convention, Amsterdam, The Netherlands (2008).
[4]V. Välimäki, J. Pakarinen, C. Erkut, and M. Karjalainen, "Discrete-time modelling of musical instruments," Rep. Prog. Phys. **69**, 1–78 (2006).
[5]*Five Lectures on the Acoustics of the Piano*, edited by A. Askenfelt (Royal Swedish Academy of Music, Stockholm, Sweden, 1990); http://www.speech.kth.se/music/5_lectures/ (Last viewed October, 2008).
[6]G. De Poli, F. Campetella, and G. Borin, "Pedal resonance effect simulation device for digital pianos," U.S. Patent No. 5,744,743 (1998).
[7]S. Zambon, H.-M. Lehtonen, and B. Bank, "Simulation of piano sustain-pedal effect by parallel second-order filters," in Proceedings of the 11th International Conference on Digital Audio Effects (DAFx-08) Espoo, Finland (2008), pp. 199–204; http://www.acoustics.hut.fi/dafx08/papers/dafx08_37.pdf (Last viewed March, 2009).
[8]Companion web page of the article, http://www.acoustics.hut.fi/go/jasael-part-pedaling/ (Last viewed May, 2009).
[9]M. Karjalainen, P. Antsalo, A. Mäkivirta, T. Peltonen, and V. Välimäki, "Estimation of modal decay parameters from noisy response measurements," J. Audio Eng. Soc. **50**, 867–878 (2002).
[10]G. Weinrech, "Coupled piano tones," J. Acoust. Soc. Am. **62**, 1474–1484 (1977).
[11]K. A. Legge and N. H. Fletcher, "Nonlinear generation of missing modes on a vibrating string," J. Acoust. Soc. Am. **76**, 5–12 (1984).
[12]E. Rank and G. Kubin, "A waveguide model for slapbass synthesis," in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Munich, Germany (1997), pp. 443–446.
[13]A. Askenfelt and E. V. Jansson, "From touch to string vibrations. III: String motion and spectra," J. Acoust. Soc. Am. **93**, 2181–2196 (1993).

Sound absorption of porous metals at high sound pressure levels

Xiaolin Wang, Feng Peng, and Baojun Chang
Key Laboratory of Noise and Vibration Research, National Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China
wangxl@mail.ioa.ac.cn, pengfeng@mail.ioa.ac.cn, changbaojun@mail.ioa.ac.cn

Abstract: This paper is a study about sound absorption properties of porous metals at high sound pressure levels. A method of deriving the nonlinear static flow resistance for highly porous fibrous metals is proposed by solving Oseen's equation to take account of the inertia effect, validated by experiments of airflow measurement. In order to predict nonlinear sound absorbing performance of a finite thickness porous metal layer, a numerical method is employed, verified by sound absorption measurement in an impedance tube. Accordingly, the effects of the nonlinear coefficient on the porous metal sound absorption are investigated.

PACS numbers: 43.25.Dc, 43.25.Ba, 43.50.Gf, 43.55.Ev [MH]
Date Received: February 21, 2009 **Date Accepted:** June 8, 2009

1. Introduction

Because of advantages in heat resistance and specific strength, porous metals have potential applications in noise control under extreme conditions, such as high sound pressure levels (SPLs) and high temperatures. Those applications include aircraft engine liners and combustion chambers for rocket engines. It was reported that compressed metallic foams in a liner can enhance the broadband characteristics of noise reduction.[1]

It is well known that at high SPLs, the airflow velocity inside porous materials tend to be so large that the flow resistance no longer follows Darcy's law. This significantly affects the sound absorbing performance of porous materials. Zorumski and Parrot[2] found that the acoustic impedance is airflow velocity dependent at high sound-intensities. Kuntz and Blackstock[3] showed that excess attenuation can be caused by the flow resistivity deviation from Darcy's law. A Forchheimer type nonlinearity was introduced by Nelson[4] to accommodate acoustic signals. Wilson *et al.*[5] further extended the Forchheimer nonlinearity to the resistive damping and mass reactance terms in complex density. This model was later modified by McIntosh and Lambert.[6] A new set of parameters in the flow resistance is thus introduced. It is noted that the nonlinear thermal effect is insignificant in porous materials.

Wilson and Lambert *et al.*[5,7] also introduced a numerical as well as a wave perturbation method to solve sound propagation inside porous materials, especially attenuation and surface admittance at high intensities. With a different set of parameters, such as viscous and thermal characteristic lengths but still based on Forchheimer's correction to Darcy's law, Umnova *et al.*[8] introduced another model to deal with high-intensity sound propagation through hard-backed rigid-porous layers. In this model, changes in wave amplitude due to the nonlinearity are assumed to be small when the distance of wave traveled is comparable to the wavelength.

In this article, the static nonlinear coefficient in the Forchheimer relation is deduced for highly porous fibrous metals. A numerical method is also introduced to calculate one-dimensional sound propagation in porous metals. The porous frame is assumed to be rigid. Experiments were conducted to verify the calculated results both for fibrous and lotus-type porous metals. The former metals consist of sintered metallic fibers. The latter (also called as gasars) contain unidirectionally aligned cylindrical pores.[9] This is a preliminary study in order to explore possible applications at high SPLs.

2. Forchheimer parameter for highly porous fibrous metals

Forchheimer's experimental law presents a simple relation that the static flow resistivity σ is linear with the cross-sectional average airflow velocity u inside a bulk porous material:

$$\sigma = \sigma_0(1 + \xi|u|), \tag{1}$$

where ξ is the Forchheimer parameter, or the nonlinear coefficient in terms of relations between the pressure gradient and the velocity.

Here the authors intend to deduce this parameter for highly porous fibrous metals, which could as well have good sound absorption performance at linear conditions, i.e., at low SPLs. The metal fibers are assumed to be laid out in parallel to each other, all perpendicular to the flow direction. The average diameter of the fibers is $2a$, and the average distance between the centers of two neighboring fibers is $2R$. A unit cell is hypothetically enveloped by a cylindrical fluid surface, approximately of radius R.

The airflow problem is assumed to be at low Reynolds numbers. The flow resistivity is obtained by solving Oseen's equation, a linearized Navier–Stokes equation that the nonlinear convection or inertia term is partially retained.[10]

$$\rho \mathbf{U}_0 \cdot \nabla \mathbf{u} = -\nabla \mathbf{p} + \mu \nabla^2 \mathbf{u}, \tag{2}$$

where ρ is the air density, μ is the dynamic viscosity of the airflow, p is the pressure, $\boldsymbol{u}$ is the airflow velocity inside the pores, and $\mathbf{U}_0$ is the ambient airflow velocity. The boundary conditions are that of no-slip condition on the fiber cylinder surface and zero shear stress on the hypothetical fluid surface boundary of each cell. After solving the equation by a method similar to matched asymptotic expansions, a semi-empirical solution is obtained:

$$\sigma \approx \frac{2.9\mu\phi(1-\phi)}{a^2}\left(1 + \frac{1}{5}\,\mathrm{Re}\right), \tag{3}$$

where the porosity $\phi = 1 - (a/R)^2$, and Re is the Reynolds number based on the diameter of the fiber.

3. Solution to a porous metal layer with finite thickness

Following the one-dimensional acoustical equations of motion and continuity given by Wilson and McIntosh *et al.*,[5,6] and taking $\exp(i\omega t)$ time dependence, the equations describing sound wave propagating inside porous materials can be written in the following form:

$$-\frac{dp}{dx} = i\omega\rho v, \tag{4}$$

$$-\frac{dv}{dx} = \frac{i\omega}{K}p, \tag{5}$$

where ω is angular frequency, $u = \phi v$, and ρ and K are complex density and bulk modulus.

$$\rho = \rho_p + \frac{8\xi\sigma_0\phi|u|}{3\pi i\omega}, \tag{6}$$

$$K = K_p, \tag{7}$$

where ρ_p and K_p are the linear complex density and liner bulk modulus, respectively. Now if a porous metal layer with a thickness of L_p backed by a cavity of depth L_c, then from the above four equations the authors can arrive at a second order nonlinear ordinary differential equation:

Table 1. Parameters of fibrous porous samples.

Sample No.	1	2	3	4	5	6
Radius a (μm)	4	6	10	19.7	19	25
Thickness L_p (mm)	6.9	15.9	12.8	13.5	10	9
Porosity ϕ (%)	95.75	92.98	96.43	96	92.8	95.1

$$\frac{d^2u}{dx^2} + k_p^2(\omega)u + C|u|u = 0, \quad C = \frac{8\omega\xi\sigma_0\phi}{3\pi i K_p} = \frac{8\xi\sigma_0 k_p}{3\pi i Z_p}, \tag{8}$$

where k_p and Z_p are the linear complex wave number and characteristic impedance of the porous material, respectively. The boundary conditions are at the incident surface the velocity and pressure must be continuous, while at the backing the surface impedance is assumed given. The above equation is first non-dimensionalized with a characteristic velocity, and then linearized and solved with a finite difference method.

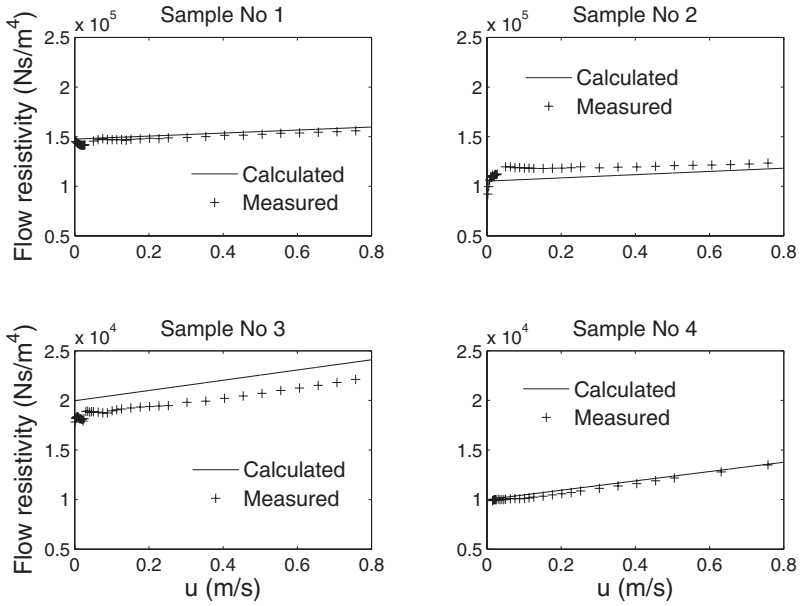
4. Results and discussion

Both fibrous and lotus-type porous metals are used for the test. Six fibrous metal samples, Nos. 1–6, and two lotus-type porous metal samples, Nos. 7 and 8, were tested. The airflow resistivities of all these samples were measured according to the ASTM C522-73 standard. The surface impedances and sound absorption coefficients were measured at various incident SPLs by the transfer function method in an impedance tube, according to the ISO 10534-2 standard.

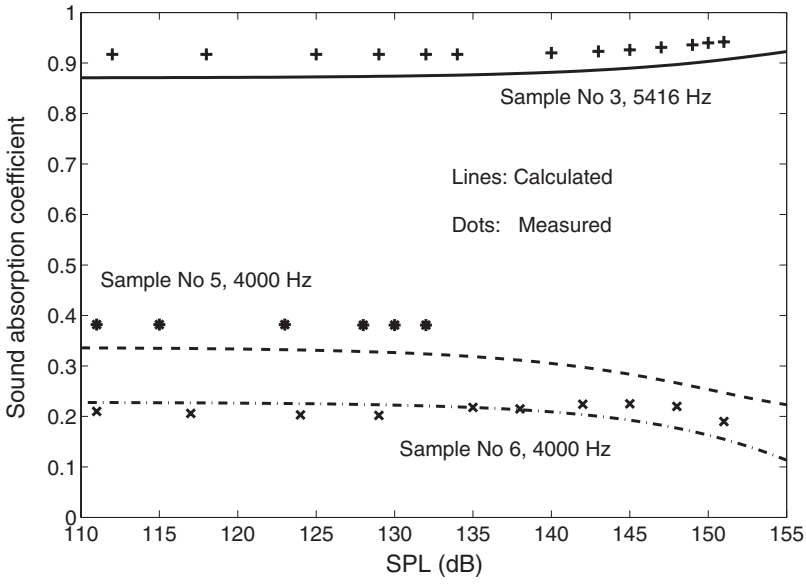
Parameters of Sample Nos. 1–6 are listed in Table 1. The static flow resistivities of Sample Nos. 1–4 were measured as in Fig. 1(a). Then the sound absorption of sample Nos. 3, 5, and 6 were measured at high SPLs as in Fig. 1(b). The static flow resistivities of Sample Nos. 7 and 8 were also measured, but both σ_0 and ξ were obtained by linear fit from experiments. These parameters were then used to predict the surface impedance and sound absorption coefficient at high SPLs, as shown in Fig. 2. The effect of the nonlinear coefficient ξ on the sound absorption is then examined in Fig. 3 by introducing two dimensionless parameters.

Figure 1(a) shows that the linear static resistivity σ_0 decreases while the nonlinear coefficient ξ increases with fiber diameter, when the porosity keeps roughly at the same order. It was also found that the cross-sectional shape of the fiber has little effect on the calculation of the resistivity, if a hydraulic radius is used for the average radius calculation. The hydraulic radius is defined as twice the cross-sectional area divided by the circumference of the fiber. Figure 1 shows that the calculation agrees fairly well with experiments. This agreement validates the perturbation solution to Oseen's equation at low Reynolds numbers, where the flow is in the transition regime between laminar and turbulent flows. Therefore, this agreement indicates that, at high SPLs, part of the airflow resistance is spent on the generation of eddies that are not yet developed into turbulent. This in a way also verifies the well known explanation that the nonlinearities are resulted from the inertia term in the Navier–Stokes equation, i.e., from drag instead of turbulence.[5]

Two typical absorption curves of porous metals at resonance frequencies at high SPLs are shown in Fig. 2. At these frequencies, sound absorption coefficient is usually large, but for one type the coefficient monotonically drops with increasing SPLs as in Fig. 2(a). For another type, the coefficient first increases, comes to a maximum, and then rapidly drops as in Fig. 2(b). The two calculated curves follow almost exactly with the experimental data. Wilson *et al.*[5] also calculated these trends, but their experimental data were unable to verify the trends because the range of their data was limited below SPL 140 dB. When the surface impedance instead of admittance is drawn, it will be clearly shown that if the acoustic resistance is larger than $\rho_0 c_0$ the absorption coefficient will follow the trend of Fig. 2(a); if with increasing SPLs the acoustic resistance is less than $\rho_0 c_0$ first and then monotonically climbs over $\rho_0 c_0$, the absorption trend will be like Fig. 2(b).



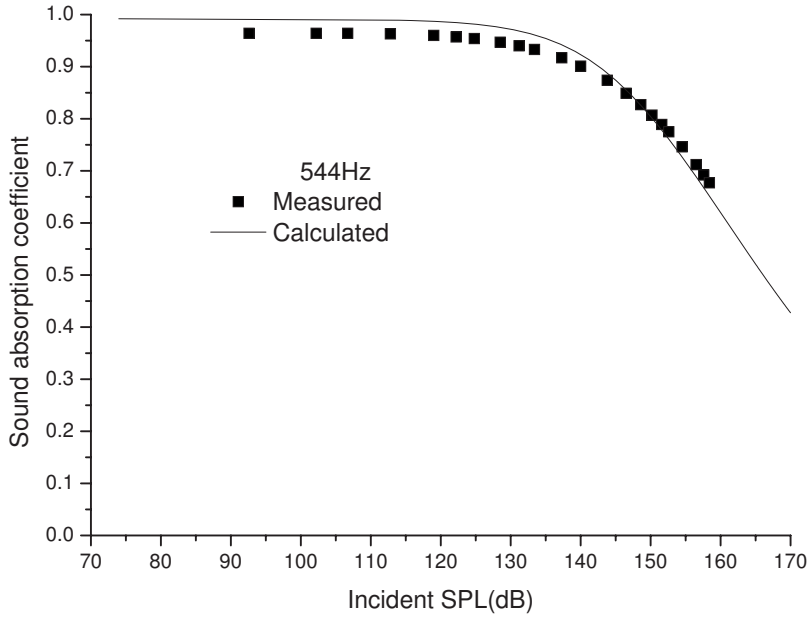
(a)



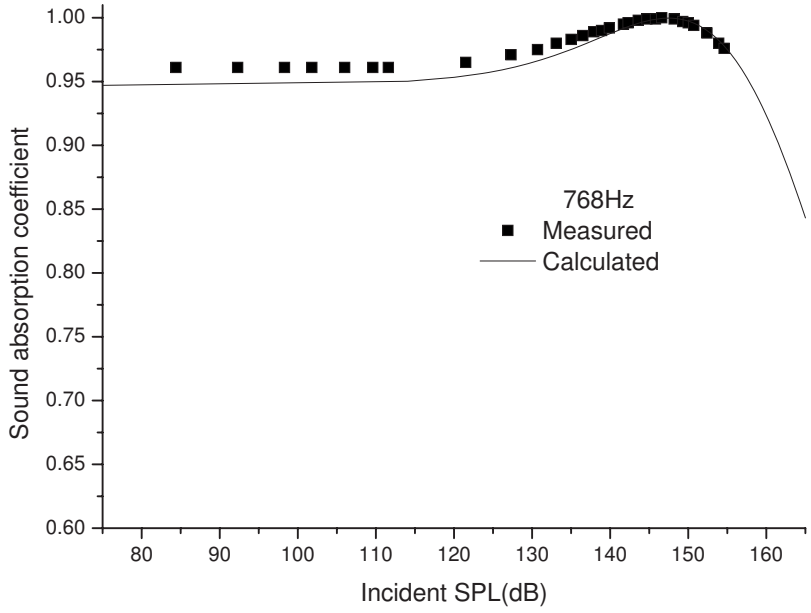
(b)

Fig. 1. Fibrous porous metals: (a) static flow resistivity (calculated ξ=0.047, 0.069, 0.12, 0.136 for Sample Nos. 1–4, respectively) and (b) sound absorption of bulk samples with rigid backing at high SPLs, both at a resonance frequency (No. 3) and not at resonance frequencies (Nos. 5 and 6).

The relationship between the nonlinearity, frequency, sound absorption coefficient, and incident SPL is shown in Fig. 3. In order to exclude the effects of porous layer thickness, the thickness is taken as close to a characteristic length, $\Lambda \equiv c_0\rho_0/\phi\sigma_0$, with which the layer has an absorption relationship commensurate with a layer of infinite thickness. No cavity is assumed so that the cavity effect is also excluded. The nonlinearity is represented by a combination of the



(a)



(b)

Fig. 2. Lotus-type porous metals: sound absorption of bulk samples with a cavity in front of rigid backing. (a) No. 7: σ_0=29 981 Ns/m^4, ξ=1.93 s/m, L_p=14.5 mm, L_c=41 mm, (b) No. 8: σ_0=13901 Ns/m^4, ξ=0.81 s/m, L_p =15.2 mm, and L_c=44 mm.

Forchheimer parameter and the incident sound pressure: a dimensionless parameter $\gamma \equiv \xi|p_i|/\rho_0 c_0$, where $|p_i|$ is the modulus of the complex incident sound pressure, and $\rho_0 c_0$ is the characteristic impedance of air. The frequency is represented by another dimensionless param-

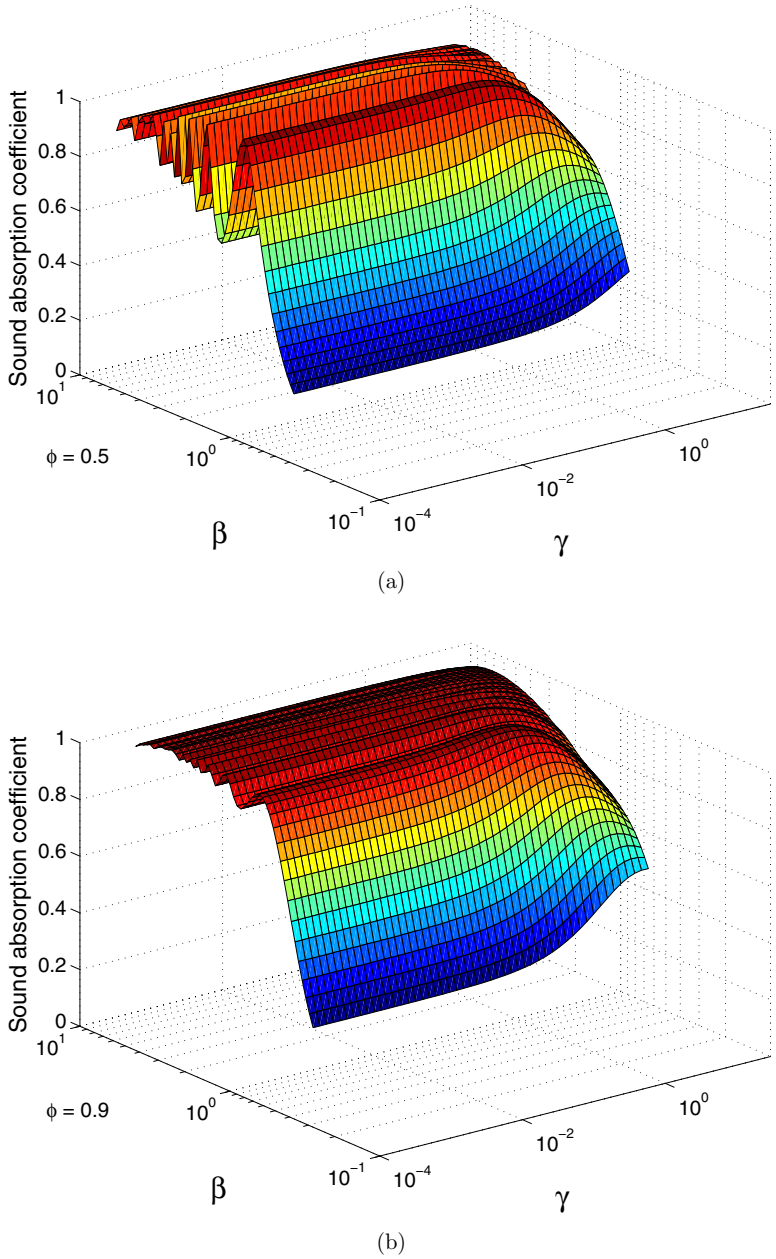


Fig. 3. (Color online) Sound absorption coefficient of porous layers with rigid backing and the nonlinear effects at different porosities. $L_p/\Lambda=0.86$. (a) $\phi=0.5$ and (b) $\phi=0.9$.

eter $\beta \equiv \sqrt{\omega\rho_0/\phi\sigma_0}$, which characterizes the ratio of the pore size to the viscous boundary thickness inside the pores. Figure 3 shows that with increasing porosity the sound absorption is less oscillating with β. The two types of sound absorption trends in Fig. 2 also appear here in Fig. 3 as the sound absorption varies with γ, at constant β of resonance frequencies. Figure 3 also indicates that the greater the porosity, the more likely the sound absorption will keep to be consistently large in a wide range of frequencies and SPLs.

5. Conclusion

This study shows that the linear static resistivity σ_0 decreases while the nonlinear coefficient or the Forchheimer parameter ξ increases with fiber diameter, when the porosity keeps roughly at the same order. The trend of sound absorption coefficient at a resonance frequency under high intensities is largely reflected by the ratio of acoustic resistance to the characteristic impedance of air, depending on whether this ratio is greater or less than 1, and how this ratio changes with the increasing SPLs. Both theoretical calculations for the Forchheimer parameter for highly porous fibrous metals and the solution to porous metal layers are in good agreement with experiments.

Acknowledgments

This work is supported by the National Basic Research Program of China (No. 2006CB601204) and the Natural Science Foundation of China (No. 1077415). The authors wish to thank H. P. Tang, J. L. Zhu, and Q. B. Ao of the Northwest Institute for Non-ferrous Metal research, and Y. X. Li, Y. Liu, and J. Wan of Tsinghua University for providing porous metal samples.

References and links

[1]P. Nordin and S. L. Sarin, "Development of new linear technology for application in hot stream areas of aero-engines," AIAA Paper No. 2004-3033.
[2]W. E. Zorumski and T. L. Parrott, "Nonlinear acoustic theory for rigid porous materials," NASA, Washington, D.C. 20546, Report No. TN D-6196, 1971.
[3]H. L. Kuntz and D. T. Blackstock, "Attenuation of intense sinusoidal waves in air-saturated bulk porous materials," J. Acoust. Soc. Am. **81**, 1723–1731 (1987).
[4]D. A. Nelson, "Interaction of finite-amplitude sound with air-filled porous materials," NASA Lewis Research Center, Cleveland, OH, 44135, Report No. CR 174885, 1985.
[5]D. K. Wilson, J. D. McIntosh, and F. R. Lambert, "Forchheimer-type nonlinearities for high-intensity propagation of pure tones in air-saturated porous media," J. Acoust. Soc. Am. **84**, 350–359 (1988).
[6]J. D. McIntosh and R. F. Lambert, "Nonlinear wave propagation through rigid porous materials. I: Nonlinear parameterization and numerical solutions," J. Acoust. Soc. Am. **88**, 1939–1949 (1990).
[7]R. F. Lambert and J. D. McIntosh, "Nonlinear wave propagation through rigid porous materials. II: Approximate analytical solutions," J. Acoust. Soc. Am. **88**, 1950–1955 (1990).
[8]O. Umnova, K. Attenborough, E. Standley, and A. Cummings, "Behavior of rigid-porous layers at high levels of continous acoustic excitation: theory and experiment," J. Acoust. Soc. Am. **114**, 1346–1356 (2003).
[9]Y. Liu, Y. X. Li, J. Wan, and H. W. Zhang, "Evaluation of porosity in lotus-type porous magnesium fabricated by metal/gas eutectic unidirectional solidification," Mater. Sci. Eng., A **402**, 47–54 (2005).
[10]M. Van Dyke, *Perturbation Methods in Fluid Mechanics* (The Parabolic, Stanford, 1975).

 [DOI: 10.1121/1.3175882] Published Online 21 July 2009

Limiting unwanted cues via random rove applied to the yes-no and multiple-alternative forced choice paradigms

Huanping Dai
Department of Speech, Language, and Hearing Sciences, The University of Arizona, Tucson, Arizona 85721
hdai@email.arizona.edu

Gerald Kidd, Jr.
Department of Speech, Language and Hearing Sciences and Hearing Research Center, Boston University, Boston, Massachusetts 02215
gkidd@bu.edu

Abstract: When a random rove is used in a perceptual task to control the influence of an unwanted cue that may confound the decision strategy of primary interest, the effectiveness of the rove is determined by its range. Green [*Profile Analysis* (Oxford University Press, Oxford, 1988)] provided a formula which allows experimenters to determine the roving range required to ensure that the listeners relying on the unwanted cue cannot exceed a predefined percentage of correct responses in a two-interval, two-alternative forced-choice experiment. Here, Green's analysis is extended to the yes-no and m-alternative, forced-choice paradigms ($m > 2$).

PACS numbers: 43.66.Ba, 43.66.Fe, 43.66.Dc, 43.66.Hg [QJF]
Date Received: April 10, 2009 **Date Accepted:** June 19, 2009

1. Introduction

In designing an auditory detection or discrimination experiment, the researcher often faces the possibility that multiple decision strategies, or perceptual cues, are available to the listeners in performing the task. For example, in a spectral-shape discrimination experiment, the primary interest of the researcher is the listeners' ability to detect a change in the relative level of different frequency components. However, a change in the relative level is always accompanied by changes in the absolute level of certain frequency components. As a result, the task can also be performed based on the absolute level of a single component, without having to monitor the relative level of different frequency components at all. Unless the influence of the change in absolute level on the listener's decision is under control, the outcome of such an experiment will not have any definitive implication on the listeners' ability to discriminate spectral shapes. Any cue other than the cue of primary interest is denoted as an unwanted cue in this paper. (Thus, in the above example the absolute-level cue is an unwanted cue). One approach to limiting the influence of an unwanted cue is to apply random rove to the stimulus dimension associated with that cue, de-correlating the cue from the primary cue. For example, by randomly varying the absolute overall stimulus level upon each stimulus presentation independent of changes made to spectral shape, any change in absolute level will be an unreliable indication of a change in spectral shape. For this reason, random rove has been an essential component in studies of spectral-shape discrimination (e.g., Spiegel *et al.*, 1981; Green, 1988). Aside from its role in studies of spectral-shape discrimination, the randomization technique has had a long history in auditory research, and has been applied in a wide range of studies and tasks to control the influence of various unwanted cues, including frequency discrimination or pitch perception (e.g., Henning, 1966; Moore and Glasberg, 1989; Emmerich *et al.*, 1989; Dai *et al.*, 1994; Shackleton and Carlyon, 1994; Carlyon and Shackleton, 1994; Moore *et al.*, 2006; Micheyl and Oxenham, 2007; Oxenham *et al.*, 2009), binaural hearing (e.g., Koehnke and Colburn, 1987;

Wightman and Kistler, 1999; Zahorik, 2002; Stellmack *et al.*, 2006; Fan *et al.*, 2008), and cases where multiple cues were controlled in a single task (e.g., Richards *et al.*, 1989; Lutfi, 2001).

The extent to which an unwanted cue is controlled by a random rove is directly determined by the range of the rove. For the two-interval, two-alternative forced-choice (2I-2AFC) paradigm, Green (1988) (pp. 19–21) provided a quantitative analysis of the limit that the random rove imposed on the discrimination performance that can be achieved on the basis of an unwanted cue. Specifically, for a given size of the change in the variable associated with the unwanted cue, Δ, and a range of rove, R, the upper limit of percent correct based on the unwanted cue is given by $\mathrm{PC}(\Delta)=0.5+\Delta/R-0.5(\Delta/R)^2$. Green's formula can be used as a guideline in several ways for applying the randomization technique. For example, suppose the experimenter in a spectral-shape discrimination experiment applies a rove of $R=20$ dB to limit an unwanted absolute-level cue of $\Delta=1$ dB. According to Green's formula, the unwanted cue alone can produce a percent correct of PC=54.9% at the best. In another example, suppose the experimenter wants to determine the range of rove needed to limit a 1-dB unwanted cue to a percent correct of PC=52%. According to Green's formula, the rove must have a range of $R=\Delta/[1-\sqrt{2(1-\mathrm{PC})}]=49.5$ dB to accomplish that objective.

Green's formula was derived specifically for applications with a 2I–2AFC paradigm; to the authors' best knowledge, it has not been extended to other psychophysical paradigms. Because different paradigms have their own special merits (Macmillan and Creelman, 2005), there are occasions when the experimenter wants to apply the randomization technique under other paradigms than the 2I–2AFC option. For example, Versfeld and Houstma (1991, 1995) applied random rove to spectral-shape discrimination under a 3AFC paradigm. Therefore, it is useful to provide a theoretical analysis of the limit imposed by randomization on the performance that can be achieved based on an unwanted cue in other paradigms than 2–I2AFC. The primary purpose of this note is to extend the analysis by Green (1988) to the yes-no and general multiple-alternative, forced-choice (mAFC) paradigms. The analysis addresses random roves applied to both continuous and discrete variables.

2. Analysis and discussion

2.1 Random rove applied to continuous variables

The following analysis is limited to randomizations based on uniform distributions which have been used in most applications of the randomization technique in auditory-perception research. Among different distributions of a fixed range, the uniform distribution has been shown to produce the lowest limit on percent correct (Dai, 2008). Let x be the stimulus variable associated with the unwanted cue, upon which randomization is applied. For the standard-alone condition, the distribution has a probability density function of $y_n(x)=f(x)=1/R$, defined within the range $R>x>0$, and is zero elsewhere. For the standard-plus-signal condition, it is $y_S(x)=f(x-\Delta)$, defined within $R+\Delta>x>\Delta$, and is zero elsewhere.

In the yes-no task, the optimum decision rule based on likelihood ratio is to respond "yes" if the magnitude of the observation is greater than a fixed criterion. The criterion is unbiased if it corresponds to a likelihood ratio of unity (assuming that equal number of no-signal and signal trials are included). The proportion of correct responses is identical to the proportion of yes responses over the trials when the signal is presented. Given this decision rule and the two density functions $f(x)$ and $f(x-\Delta)$, the proportion of correct and unbiased responses for Δ can be expressed as a weighted sum of two PCs:

$$\mathrm{PC}_{\text{yes-no}}(\Delta)=\left(1-\frac{\Delta}{R}\right)\mathrm{PC}_0+\frac{\Delta}{R}\mathrm{PC}_1,$$

in which $\mathrm{PC}_0=50\%$ represents the chance performance when the observation or sample of x falls into the overlapping region (i.e., where the likelihood ratio is unity) of the two distributions, and $\mathrm{PC}_1=100\%$ represents the perfect performance when the sample falls into the nonoverlapping region (i.e., where likelihood ratio is infinity) of the shifted distribution. Combin-

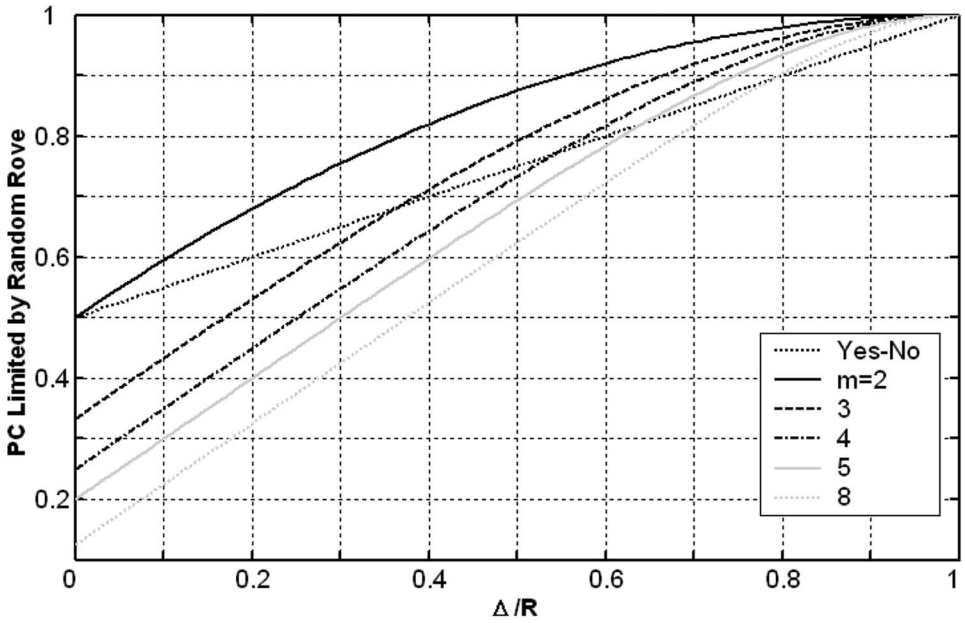


Fig. 1. Percentage of correct responses (PC) based on unwanted cues (Δ), which is limited by random rove implemented with continuous uniform distributions, as a function of the ratio between Δ and the range of the rove (R). Different curves represent the results computed for different psychophysical paradigms, including the yes-no (dark dotted line) and mAFC paradigms with $m=2-8$ (as indicated by the legends). The intercepts of the curves on the y axis (i.e., $\Delta=0$) represent the chance performances for the specific paradigms.

ing the above expressions, we obtain

$$\mathrm{PC}_{\text{yes-no}}(\Delta)=\frac{1}{2}+\frac{1}{2}\frac{\Delta}{R}. \tag{1}$$

This function is plotted as the dark dotted line in Fig. 1.

In a mAFC paradigm, in which the random rove is implemented with a uniform distribution, the optimum decision rule is to respond to the interval in which the magnitude of the observation variable is greater than that in any of the other $m-1$ intervals. Given that the sample drawn from the standard-plus-signal distribution equals $x=\eta$, the probability that every sample drawn from the $m-1$ standard-alone distributions is less than η would be $[\int_{-\infty}^{\eta} y_N(\zeta)d\zeta]^{m-1}$. This probability integrated over *all* possible values of η, therefore, would be the PC value, which can be expressed generally as (Green and Swets, 1966) (p. 47):

$$\mathrm{PC}_m(\Delta)=\int_{-\infty}^{\infty} y_S(\eta)\left[\int_{-\infty}^{\eta} y_N(\zeta)d\zeta\right]^{m-1} d\eta. \tag{2}$$

By plugging in the above-defined probability density functions into Eq. (2), PC can be further expressed as

$$\mathrm{PC}_m(\Delta)=\int_{\Delta}^{R+\Delta} f(\eta-\Delta)\left[\int_{0}^{\eta} f(\zeta)d\zeta\right]^{m-1} d\eta. \tag{3}$$

Substituting the expression for $f(x)$ into Eq. (3) and integrating, we have

$$\mathrm{PC}_m(\Delta)=\frac{1}{m}\left[1-\left(\frac{\Delta}{R}\right)^m\right]+\frac{\Delta}{R}. \tag{4}$$

The first term of Eq. (4) represents the performance when the sample falls into the overlapping region of the shifted signal distribution and the $m-1$ un-shifted non-signal distributions. The second term represents the perfect performance when the sample falls into the non-overlapping region of the shifted signal distribution. This function is plotted in Fig. 1 for $m=2$, 3, 4, 5, and 8.

Figure 1 shows for different paradigms the limit imposed by a uniformly distributed random rove on the percent-correct performance that can be achieved based on an unwanted cue. For each paradigm, the proportion of correct responses (PC) attainable using the unwanted cue is calculated using Eq. (1) (for the yes-no paradigm) or Eq. (4) (for mAFC paradigms), and is plotted as a function of the ratio between the size of an unwanted cue (Δ) and the range of the rove (R). Given that a uniform distribution will be used to limit the unwanted cue in the yes-no or mAFC tasks, the researcher can use Eq. (1) or (4) to determine how big a range is sufficient to keep performance based on the cue from reaching a certain value. For relatively small ratios of Δ/R, mAFC paradigms with greater m values produced lower limits on PC, largely due to the lower baseline or chance performance level for these paradigms. While the yes-no and 2I–2AFC paradigms share the same chance performance at 50%, random rove implemented with the yes-no paradigm produced percent correct values below those of the 2I–2AFC paradigm at all values of Δ/R. As the ratio Δ/R increases, random rove implemented with the yes-no paradigm becomes increasingly more effective relative to that implemented with the mAFC paradigms. Beyond about $\Delta/R=0.7$, the yes-no paradigm produces the lowest percent correct values than all the mAFC paradigms shown.

2.2 Random rove applied to discrete variables

Thus far, the analysis has been carried out for continuous variables. However, in experimental applications of the randomization technique, the experimenter may choose to draw samples from a discrete distribution. Although discrete distributions can be approximated by continuous ones when their probability mass functions have a very large number of bins, the results obtained with continuous and discrete distributions become clearly different when the bins are sparse. Thus, in general, randomization applied to discrete variables should not be treated as a special case of that applied to continuous variables. In particular, Eqs. (1) and (4) will lead to sizable error when the number of bins is small. In practice, a small number of bins in discrete distributions can happen when the range of randomization (R) is comparable to the size of the signal (Δ), which may result from either a severely limited range to apply randomization (as with hearing impaired ears) or an unusually large signal size (as with an extremely difficult task, e.g., Dai and Green, 1992). Therefore, we need to derive for discrete variables the equations relating PC to Δ and R, which is the purpose of this section.

For analyzing random rove applied to discrete variables, we define the probability mass function of the random rove as having n bins spaced at equal inter-bin intervals of $b=R/(n-1)$, with each bin representing a probability of $p=1/n$. Note that the amount of shift of the distribution caused by the unwanted cue, Δ, is allowed to take only integer numbers of the inter-bin intervals; other values of Δ will produce no overlap of bins between the distributions, leading to a percent correct of 100%, and rendering the application of random rove useless. With a shift of Δ between the two distributions for the non-signal and signal events, the number of bins that do not overlap will be $\Delta/b=(n-1)\Delta/R$, which amounts to a probability of $p\Delta/b=[(n-1)/n]\Delta/R$. Note that for continuous distributions the probability of the non-overlapping portion is simply Δ/R. Thus, the expressions for the limits on PC imposed by random rove for discrete variables can be obtained by replacing Δ/R with $[(n-1)/n]\Delta/R$ in Eqs. (1) and (4). We have

$$\mathrm{PC}_{\text{yes-no}}^{\text{discrete}}(\Delta)=\frac{1}{2}+\frac{1}{2}\frac{n-1}{n}\frac{\Delta}{R}, \tag{5}$$

for the yes-no paradigm, and

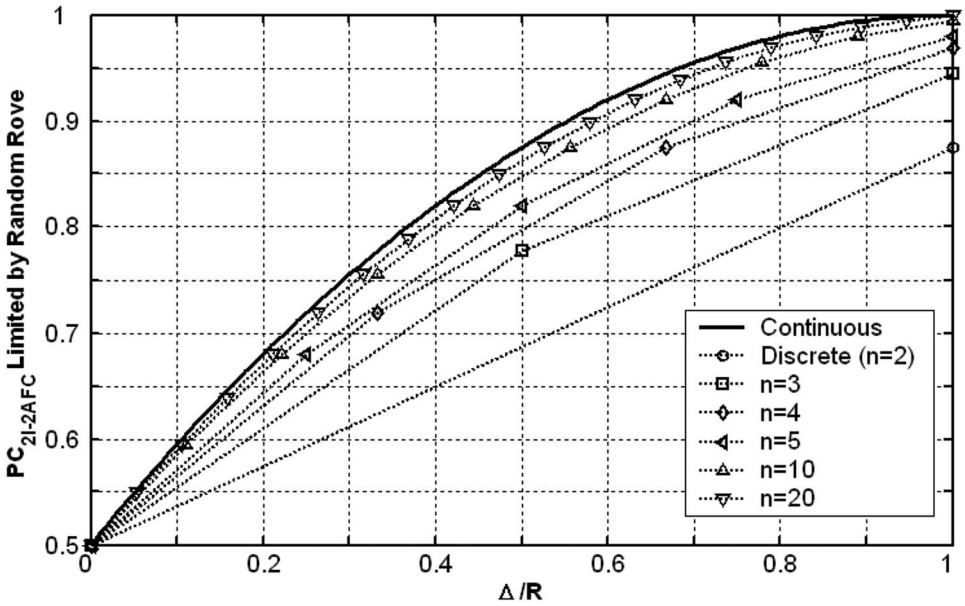


Fig. 2. Same format as in Fig. 1, but for the 2I–2AFC paradigm only. The symbols represent the percent correct values limited by random rove implemented with discrete uniform distributions with $n=2$ (circles), 3 (squares), 4 (diamonds), 5 (right triangles), 10 (up triangles), and 20 (down triangles) bins. The solid line represents the function obtained with continuous uniform distributions in Fig. 1, which is replotted her for comparison.

$$\mathrm{PC}_m^{\mathrm{discrete}}(\Delta) = \frac{1}{m}\left[1 - \left(\frac{n-1}{n}\frac{\Delta}{R}\right)^m\right] + \frac{n-1}{n}\frac{\Delta}{R}, \tag{6}$$

for the mAFC paradigms. These results may be interpreted as if, by implementing the random rove with discrete instead of continuous distributions, the size of the unwanted cue is effectively reduced by a factor of $(n-1)/n$.

To illustrate the different outcomes of random rove applied to discrete and continuous variables, the limit on the percent-correct performance for the 2I–2AFC paradigm is computed as a function of Δ/R using Eq. (6), for discrete variables with several values of the number of bins. Figure 2 presents the results obtained with $n=2$ (circles), 3 (squares), 4 (diamonds), 5 (right triangle), 10 (up triangle), and 20 (down triangle) bins. For comparison, the solid dark curve from Fig. 1, which represents the result obtained using a continuous distribution with the 2I–2AFC paradigm, is re-plotted here. A general summary of the results is that, for all possible values of Δ/R where the discrete distributions *are* available, random rove leads to lower percent correct values, thus is more effective, when implemented with discrete than with continuous distributions. The advantage of discrete distributions is most pronounced for distributions with five or fewer bins. For example, with $n=2$ and a shift of $\Delta=R$, the percent correct is 100% with a continuous distribution, but is limited to PC=87.5% with a discrete distribution. With continuous distributions, to limit the unwanted cue to the same PC=87.5% would require a rove range twice the size of Δ, i.e., $\Delta/R=0.5$. Thus, when $\Delta=R$, implementing random rove with the discrete distribution effectively reduces the size of the unwanted cue by half (or equivalently, effectively increases the rove range by a factor of 2) relative to that with the continuous distribution. The advantage of discrete distributions over continuous distributions diminishes as the number of bins increases. For distributions with more than $n=20$ bins, the outcomes are closely approximated by the function obtained with the continuous distribution.

Depending on the specific experimental situation, the experimenter may opt for discrete distributions with sparse bins over continuous distributions to maximize the effect of ran-

domization. This approach may prove to be a useful addition to techniques developed for experimental conditions where the range of rove is limited to not much greater than the size of the unwanted cue (e.g., Kidd and Mason, 1992; Kidd and Dai, 1993), helping us deal with special cases such as when there is a severely limited dynamic range due to hearing impairment, or when the perceptual task is extremely difficult.

Acknowledgments

We thank Chris Micheyl, Fred Wightman and one anonymous reviewer, and the associate editor, Qian-Jie Fu, for their time and very helpful comments and suggestions. This research was supported by The University of Arizona (H.D.) and NIH grants (G.K.)

References and links

Carlyon, R. P., and Shackleton, T. M. (**1994**). "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?," J. Acoust. Soc. Am. **95**, 3541–3554.
Dai, H. (**2008**). "On suppressing unwanted cues via randomization," Percept. Psychophys. **70**, 1379–1382.
Dai, H., and Green, D. M. (**1992**). "Auditory intensity perception: Successive versus simultaneous, across-channel discriminations," J. Acoust. Soc. Am. **91**, 2845–2854.
Dai, H., Nguyen, Q. T., and Green, D. M. (**1995**). "A two-filter model of frequency discrimination," Hear. Res. **85**, 109–114.
Emmerich, D. S., Ellermeier, W., and Butensky, B. (**1989**). "A reexamination of the frequency discrimination of random-amplitude tones, and a test of Henning's modified energy-detector model," J. Acoust. Soc. Am. **85**, 1653–1659.
Fan, W. L., Streeter, T. M., and Durlach, N. I. (**2008**). "Effect of spatial uncertainty of masker on masked detection for nonspeech stimuli," J. Acoust. Soc. Am. **124**, 36–39.
Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford University Press, Oxford).
Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).
Henning, B. G. (**1966**). "Frequency discrimination of random-amplitude tones," J. Acoust. Soc. Am. **39**, 336–339.
Kidd, G., Jr., and Dai, H. (**1993**). "A composite randomization technique for spectral-shape discrimination," J. Acoust. Soc. Am. **94**, 1275–1280.
Kidd, G., Jr., and Mason, C. R. (**1992**). "A new technique for measuring spectral shape discrimination," J. Acoust. Soc. Am. **91**, 2855–2864.
Koehnke, J., and Colburn, H. S. (**1987**). "Effects of roving level on binaural detection and discrimination on and off midline," J. Acoust. Soc. Am. **82**, S109.
Lutfi, R. A. (**2001**). "Auditory detection of hollowness," J. Acoust. Soc. Am. **110**, 1010–1019.
Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A User's Guide*, 2nd ed. (Erlbaum, Mahwah, NJ).
Micheyl, C., and Oxenham, A. J. (**2007**). "Across-frequency pitch discrimination interference between complex tones containing resolved harmonics," J. Acoust. Soc. Am. **121**, 1621–1631.
Moore, B. C. J., and Glasberg, B. R. (**1989**). "Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation," J. Acoust. Soc. Am. **86**, 1722–1732.
Moore, B. C. J., Glasberg, B. R., Flanagan, H. J., and Adams, J. (**2006**). "Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure," J. Acoust. Soc. Am. **119**, 480–490.
Oxenham, A. J., Micheyl, C., and Keebler, M. V. (**2009**). "Can temporal fine structure represent the fundamental frequency of unresolved harmonics?," J. Acoust. Soc. Am. **125**, 2189–2199.
Richards, V. M., Onsan, Z. A., and Green, D. M. (**1989**). "Auditory profile analysis: Potential pitch cues," Hear. Res. **39**, 27–36.
Shackleton, T. M., and Carlyon, R. P. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," J. Acoust. Soc. Am. **95**, 3529–3540.
Spiegel, M. F., Picardi, M. C., and Green, D. M. (**1981**). "Signal and masker uncertainty in intensity discrimination," J. Acoust. Soc. Am. **70**, 1015–1019.
Stellmack, M. A., Viemeister, N. F., and Byrne, A. J. (**2006**). "Discrimination of depth of sinusoidal amplitude modulation with and without roved carrier levels," J. Acoust. Soc. Am. **119**, 37–40.
Versfeld, N. J., and Houstma, A. J. M. (**1991**). "Perception of spectral changes in multi-tone complexes," Q. J. Exp. Psychol. A **43A**, 459–479.
Versfeld, N. J., and Houstma, A. J. M. (**1995**). "Discrimination of changes in the spectral shape of two-tone complexes," J. Acoust. Soc. Am. **98**, 807–816.
Wightman, F., and Kistler, D. (**1999**). "Resolution of front-back ambiguity in spatial hearing by listener and source movement," J. Acoust. Soc. Am. **105**, 2841–2853.
Zahorik, P. (**2002**). "Direct-to-reverberant energy ratio sensitivity," J. Acoust. Soc. Am. **112**, 2110–2117.

LETTERS TO THE EDITOR

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

High-frequency acoustic communications achieving high bandwidth efficiency (L)

H. C. Song,[a] W. S. Hodgkiss, and W. A. Kuperman
Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0238

T. Akal
TUBITAK-MAN, Marmara Research Center, Earth and Marine Science Research Institute, Kocaeli 41470, Turkey

M. Stevenson
Spawar Systems Center, San Diego, California 92152-5000

(Received 17 March 2009; revised 4 June 2009; accepted 7 June 2009)

A recent communications experiment was conducted in a shallow water environment at high-frequency permitting the use of a large bandwidth (11–19 kHz). This paper investigates the communication performance versus various symbol rates (or bandwidths) in terms of output signal-to-noise ratio with an assortment of constellations, illustrating a trade-off between performance and data rate. A high bandwidth efficiency of 4 bits/s Hz is demonstrated using 32 quadrature amplitude modulation with a data rate of 31.25 kbits/s over a 2.2-km range.
 [DOI: 10.1121/1.3160284]

PACS number(s): 43.60.Dh, 43.60.Gk, 43.60.Fg [NX] Pages: 561–563

I. INTRODUCTION

Underwater acoustic (UWA) communication systems typically transmit signals whose bandwidths exceed 20% of their carrier frequency, referred to as ultra-wideband (UWB) systems in radio channels.[1] Although a large bandwidth offers the potential for a high transmission rate, the available bandwidth for acoustic telemetry is very limited in UWA channels due to the relatively low carrier frequencies typically used (e.g., 15 kHz). This motivates the use of bandwidth-efficient (data rate/signal bandwidth) phase-coherent modulation with high-order constellations, and adaptive equalization is required to overcome the effects of channel delay spread.[2]

Achieving the potential of UWB, however, is quite challenging due to significant multipath in UWA channels leading to the intersymbol interference (ISI), for instance, exceeding 100 symbols at a symbol rate of 5 ksymbols/s for about 30-ms delay spread [see Fig. 1(b)]. This means that even a symbol-spaced equalizer must have at least 100 taps for a single receive element that is multiplied by the number of receiver elements, necessitating computationally formidable multi-channel equalizers. As an alternative, low-complexity multi-carrier modulation[3] [e.g., Orthogonal Frequency Division Multiplexing (OFDM)] has been proposed which eliminates the need for complex time-domain equalization of a single-carrier UWB system at the expense of bandwidth efficiency.

Recently a time reversal approach[4–6] has been introduced which exploits spatial diversity to mitigate the ISI and can provide nearly optimal performance in conjunction with channel equalization.[7] The benefit of the time reversal approach is lower computational complexity for two reasons. First, multiple receive elements are combined into a single channel and thus the complexity of successive channel equalizers remains unchanged as the number of receive elements increases. Second, the number of taps required for an equalizer is much smaller due to the temporal compression provided by time reversal combining than the number of symbols spanning the channel delay spread.

A recent communications experiment (FAF-06) explored the high-frequency regime (11–19 kHz) typical for acoustic telemetry (as compared to 3–4 kHz in previous experiments), permitting the use of a large bandwidth. The objective of this paper is to investigate the performance of time reversal communications versus various symbol rates (or bandwidths) in terms of output signal-to-noise ratio (SNR) with an assortment of constellations.

[a]Author to whom correspondence should be addressed. Electronic mail: hcsong@ucsd.edu

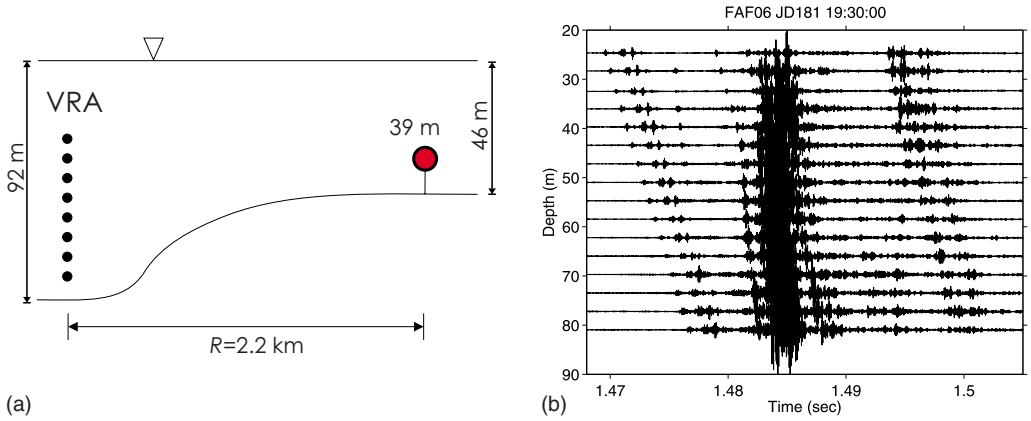


FIG. 1. (Color online) (a) Schematic of FAF-06 communications experiment. A 16-element VRA was deployed in 92-m deep water. A single element at 39-m depth is selected from a 12-element transmitter array (Topas) deployed to the seafloor in 46-m deep water at 2.2-km range. (b) An example of the channel response received by the VRA.

II. PERFORMANCE ANALYSIS

The FAF-06 experiment was carried out in shallow water June 21–July 11, 2006, off the west coast of Italy. The vertical receive array (VRA) was deployed in 92-m water and consisted of 16 elements spanning a 56.25-m aperture with 3.75-m element spacing.[8] A 12-element transmit array (Topas) with a 6.5-m aperture was deployed to the seafloor in 46-m deep water, as shown in Fig. 1(a). The distance between the two arrays was 2.2 km. During the 3-day communications experiment reported in this paper (JD 180-182), the top element at 39-m depth (Ch No. 12) was chosen as a single transmitter with the source level of 185 dB re 1 μPa. The duration of data packets was 3 s, and each packet consisted of a channel probe followed by a communications sequence. An example of the channel response is shown in Fig. 1(b), indicating about 30-ms channel delay spread.

Two aspects of this experiment are notably different than the authors's earlier experiments. First, a raised-cosine filter is employed as a shaping pulse (as opposed to a linear chirp signal used previously), where the signal bandwidth is controlled by a roll-off factor or excess bandwidth[9] (e.g., $\beta=0.5$). Second, a higher-frequency band (11–19 kHz) is utilized which enabled us to investigate the communications performance for different bandwidths. Since their interest is in high bandwidth efficiency, various constellations from binary-phase shift-keying (1 bit/symbol) up to 32 quadrature amplitude modulation (QAM) (5 bits/symbol) are examined. The receiver involves time reversal multi-channel combining followed by a single channel decision-feedback equalizer (DFE), with frequent channel updates to accommodate temporal channel variations.[10] The low complexity of the least mean squares algorithm is used for channel estimates while the recursive least squares algorithm is used for a fractionally-sampled DFE ($K=2$). This hybrid approach allows for keeping the computational complexity minimal.[10]

The resulting performance of time reversal communications is illustrated in terms of output SNR in Fig. 2 versus four distinct symbol rates ranging from 0.5 to 6.25 ksymbols/s. The corresponding bandwidths are 1 kHz ($\beta=1$), 3.75 kHz ($\beta=0.5$), 7.5 kHz ($\beta=0.5$), and 7.5 kHz ($\beta=0.2$), respectively. Note that the last two share the same bandwidth of 7.5 kHz centered at 15 kHz but with different roll-off factors. Data rates are obtained by the symbol rate multiplied by the number of bits carried by each constellation. The bandwidth efficiency (bits/s Hz) then is calculated from the data rate (bits/s) divided by the signal bandwidth (Hz).

There are two important observations. First, the extent of the performance variation for any given symbol rate is confined to about 2 dB over various constellations. Taking into account that the data are collected various times during the 3-day period, the result appears quite consistent. Interestingly, the spread is tighter at a symbol rate of 5 ksymbols/s (i.e., less than 1 dB). These results suggest that a higher-order constellation can be exploited to increase the data rate under the condition that the bit error rate (BER) corresponding to the output SNR is acceptable. Second, the performance tends to degrade with an increase in symbol rate if an average value is chosen for each symbol rate (i.e., from 29 to 23 dB), indicating a trade-off between performance and data rate. Since the transmitted power remains the same regardless of constellations and symbol rates, the transmitted symbol energy E will decrease linearly with an increase in symbol rate. It is interesting, however, that the output perfor-

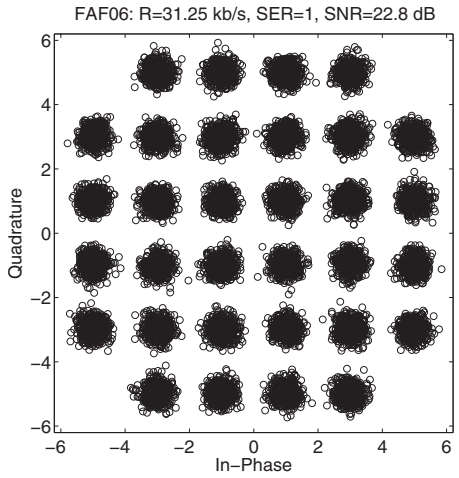


FIG. 3. Scatter plot for 32-QAM at a symbol rate of 6.25 ksymbols/s. The output SNR is 22.8 dB and the BER is 1/86 020.

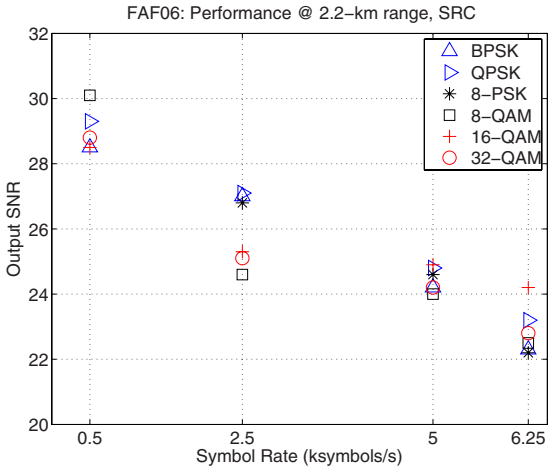


FIG. 2. (Color online) Performance of time reversal communications versus four different symbol rates with various constellations: 0.5, 2.5, 5, and 6.25 ksymbols/s. The corresponding bandwidths are 1, 3.75, 7.5, and 7.5 kHz, respectively. The high spectral efficiency of 4 bits/s Hz is achieved from 32-QAM with a data rate of 31.25 kbits/s using a 7.5-kHz bandwidth centered at 15 kHz (see Fig. 3).

mance degrades at a much slower rate. For instance, there is about 5 dB decrease in output SNR while the transmitted symbol energy decreases by 10 dB due to the increase in symbol rate from 0.5 to 5 ksymbols/s. Thus at the expense of lower performance one can achieve a high data rate by exploiting the large bandwidth available with the help of the time reversal approach which can handle significant ISI with minimal computational complexity.

Specifically, a high data rate of 31.25 ($=6.25 \times 5$) kbits/s is obtained from 32-QAM using a 7.5-kHz bandwidth centered at 15 kHz. The scatter plot is displayed in Fig. 3 with the output SNR of 22.8 dB and BER $=1/86\,020$, achieving a high bandwidth efficiency of 4 bits/s Hz. Note that at a symbol rate of 6.25 ksymbols/s the ISI spans about 200 symbols for the channel delay spread shown in Fig. 1(b). In the processing, the authors have fully exploited spatial diversity provided by the 16-element receive array (i.e., $M=16$), as shown in Fig. 2. Although not shown here, the output SNR decreases to 20 dB (about 3 dB down) when only half the array elements ($M=8$) are included in the processing.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research under Grant No. N00014-06-1-0128.

[1]S. Franz and U. Mitra, "Generalized UWB transmitted reference systems," IEEE J. Sel. Areas Commun. **24**, 780–786 (2006).
[2]D. Kilfoyle and A. Baggeroer, "The state of the art in underwater acoustic telemetry," IEEE J. Ocean. Eng. **25**, 4–27 (2000).
[3]B. Li, S. Zhou, M. Stojanovic, L. Freitag, and P. Willett, "Multicarrier communications over underwater acoustic channels with nonuniform Doppler shifts," IEEE J. Ocean. Eng. **33**, 198–209 (2008).
[4]D. Rouseff, D. Jackson, W. Fox, C. Jones, J. Ritcey, and D. Dowling, "Underwater acoustic communications by passive-phase conjugation: Theory and experimental results," IEEE J. Ocean. Eng. **26**, 821–831 (2001).
[5]T. Yang, "Temporal resolutions of time-reversed and passive-phase conjugation for underwater acoustic communications," IEEE J. Ocean. Eng. **28**, 229–245 (2003).
[6]J. Gomes, A. Silva, and S. Jesus, "Adaptive spatial combining for passive time-reversed communications," J. Acoust. Soc. Am. **124**, 1038–1053 (2008).
[7]H. C. Song and S. M. Kim, "Retrofocusing techniques in a waveguide for acoustic communications (L)," J. Acoust. Soc. Am. **121**, 3277–3279 (2007).
[8]J. D. Skinner and W. S. Hodgkiss, "A networked/autonomous receiving array system," in Proceedings of the Oceans 2008 (2008).
[9]J. Proakis, *Digital Communications* (McGraw-Hill, New York, 2001).
[10]H. C. Song, W. A. Kuperman, and W. S. Hodgkiss, "Basin-scale time reversal communications," J. Acoust. Soc. Am. **125**, 212–217 (2009).

Design of a non-traditional dynamic vibration absorber (L)

Y. L. Cheung and W. O. Wong
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong

(Received 20 August 2008; revised 15 March 2009; accepted 26 May 2009)

A non-traditional dynamic vibration absorber is proposed for the minimization of maximum vibration velocity response of a vibrating structure. Unlike the traditional damped absorber configuration, the proposed absorber has a linear viscous damper connecting the absorber mass directly to the ground instead of the main mass. Optimum parameters of the proposed absorber are derived based on the fixed-point theory for minimizing the maximum vibration velocity response of a single-degree-of-freedom system under harmonic excitation. The extent of reduction in maximum vibration velocity response of the primary system when using the traditional dynamic absorber is compared with that using the proposed one. Under the optimum tuning condition of the absorbers, it is proved analytically that the proposed absorber provides a greater reduction in maximum vibration velocity response of the primary system than the traditional absorber. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3158917]

PACS number(s): 43.40.Tm [JGM] Pages: 564–567

I. INTRODUCTION

Dynamic vibration absorber (DVA) is a passive resonator mounted on a primary system. At a pre-tuned frequency, the resonator absorbs vibration energy and contributes a sharp valley to the frequency response of the primary system. However, peaks still remain at other frequencies in the frequency response of the primary system. Therefore DVA is normally used for passive control of narrowband vibration.[1–4]

The traditional damped vibration absorber has a damper added between the absorber mass m and the primary mass M as shown in Fig. 1(a) in order to limit the vibration amplitude when it experiences lower resonance during system start-up and shut down. However, it is not possible to eliminate steady-state vibrations of the original mass after damping is added in the auxiliary mass-spring system. Optimization of the frequency and damping parameters of the traditional damped vibration absorber for minimizing the maximum amplitude response based on the fixed-point theory was well documented in the textbooks of Den Hartog[1] and Korenev and Reznikov.[2]

The traditional DVA as illustrated in Fig. 1(a) provides a cheap and convenient solution for suppressing vibration displacement amplitude of a single-degree-of-freedom system under harmonic excitation. A damped DVA in non-traditional form as shown in Fig. 1(b) has been proved by Ren[5] and Wong and Cheung[6] to be more effective than the traditional absorber in minimizing the maximum vibration displacement response of the primary system under force and ground motion excitation, respectively. In some cases where the dynamic absorber is used for isolating vibrations in vehicle suspension system,[7] reducing vibration energy of structures,[8] and reducing sound radiation from a vibrating surface,[9–13] the vibration absorber need to be designed such that the maximum vibration velocity in the whole spectrum of the primary system needs to be minimized. The optimum tuning frequency and damping of a vibration absorber are different from those values reported by Ren[5] if the maximum velocity amplitude response is required to be minimized. While the optimum tuning frequency and damping of the traditional vibration absorber for minimizing the maximum velocity response have been reported,[14,15] those of the proposed absorber cannot be found in the literature. In this paper, the tuning frequency and damping of this damped DVA of non-traditional form have been derived for minimizing maximum vibration velocity response of a single-degree-of-freedom system. The derivation of the formulas for the optimum tuning frequency and damping of the absorber was based on the fixed-point theory of Den Hartog.[1] It is proved in Sec. III that the proposed absorber provides a greater reduction in maximum velocity amplitude response of the primary system under harmonic force excitation than the traditional absorber.

II. THE TRADITIONAL DAMPED DVA

A schematic diagram of a traditional damped DVA attached to an undamped mass-spring system under sinusoidal excitation ($f=Fe^{j\omega t}$) is shown in Fig. 1(a). This vibration model is called model A in the following discussion. The non-dimensional velocity amplitude response of the primary mass M may be written as[15]

$$G_{\mathrm{A}} = \left| \frac{\dot{X}_1}{\omega_n X_{\mathrm{st}}} \right|_{\mathrm{A}} = \sqrt{\frac{(2\zeta\beta r^2)^2 + r^2(\beta^2 - r^2)^2}{(1 - r^2 + \mu r^2)^2(2\zeta\beta r)^2 + [(1 - r^2)(\beta^2 - r^2) - \mu\beta^2 r^2]^2}}, \tag{1}$$

where $X_{\mathrm{st}}=F/K$, $\omega_n=\sqrt{K/M}$, $\omega_a=\sqrt{k/m}$, $r=\omega/\omega_n$, $\beta=\omega_a/\omega_n$, and $\zeta=c/2\sqrt{mk}$.

Using the fixed-point theory,[1] the optimum tuning frequency and damping of the absorber leading to minimum vibration amplitude at resonance are written as[15]

$$\beta_{\mathrm{opt\_A}} = \frac{1}{1+\mu}\sqrt{\frac{2+\mu}{2}}, \tag{2a}$$

 0001-4966/2009/126(2)/564/4/$25.00

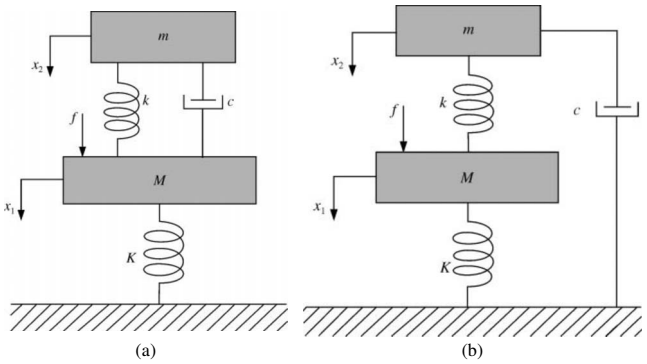


FIG. 1. A damped DVA as an auxiliary mass-spring-damper (m-k-c) system attached to a primary system (M-K): (a) traditional design of the absorber (Ref. 1) and (b) the proposed design of the absorber.

and

$$\zeta_{\text{opt\_A}} = \frac{1}{4(2+\mu)}\sqrt{\frac{\mu(24+24\mu+5\mu^2)}{1+\mu}}. \tag{2b}$$

An approximate value of the velocity amplitude ratio at resonance based on the fixed-point theory is written as

$$G_{\text{A\_opt}} = \sqrt{\frac{2+\mu}{\mu(1+\mu)}}. \tag{3}$$

Equation (3) above shows that the theoretical limit of the velocity amplitude ratio at resonance is zero when the mass ratio approaches infinity.

III. A VARIANT FORM OF THE DAMPED DVA

A variant form of the damped DVA as shown in Fig. 1(b) is called model B in the following discussion. In Fig. 1(b), the motions of the primary system and the DVA are governed by the following equations:

$$M\ddot{x}_1 = -k(x_1 - x_2) - Kx_1,$$

$$m\ddot{x}_2 = -k(x_2 - x_1) - c\dot{x}_2. \tag{4}$$

The normalized amplitude of the steady-state response of the primary mass can be derived as

$$G_{\text{B}} = \left|\frac{\dot{X}}{\omega_n X_{\text{st}}}\right| = \sqrt{\frac{(2\zeta\beta r^2)^2 + r^2(\beta^2 - r^2)^2}{(1+\mu\beta^2 - r^2)^2(2\zeta\beta r)^2 + [(1-r^2)(\beta^2 - r^2) - \mu\beta^2 r^2]^2}}. \tag{5}$$

Equation (5) may be rewritten as

$$G_{\text{B}} = \sqrt{\frac{A\zeta^2 + B}{C\zeta^2 + D}}, \tag{6}$$

where $A=4\beta^2 r^4$, $B=r^2(\beta^2-r^2)^2$, $C=4\beta^2 r^2(1+\mu\beta^2-r^2)^2$, and $D=[(1-r^2)(\beta^2-r^2)-\mu\beta^2 r^2]^2$.

Equation (5) is calculated with mass ratio $\mu=0.2$ at four different damping ratios, and the results are plotted in Fig. 2. It can be observed that there are stationary points P and Q at

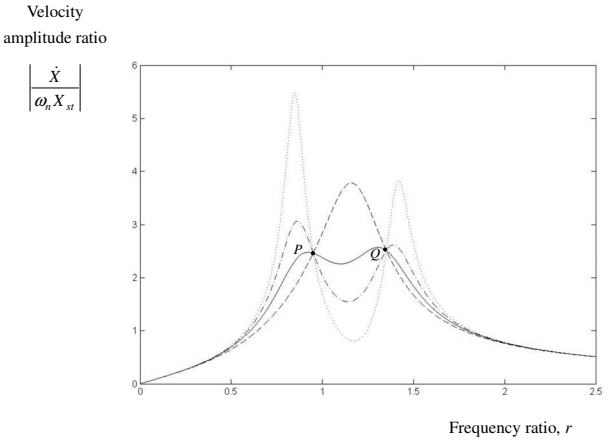


FIG. 2. Velocity amplitude ratio of the primary system of model B at different mass ratios m/M.

which the response G_{B} is independent of the damping of the absorber. At the stationary or fixed points P and Q, we may write

$$\frac{A}{C} = \frac{B}{D}, \tag{7}$$

i.e.,

$$\frac{1}{(1+\mu\beta^2 - r^2)^2} = \frac{(\beta^2 - r^2)^2}{[(1-r^2)(\beta^2 - r^2) - \mu\beta^2 r^2]^2} \tag{8a}$$

or

$$2r^4 - 2(1+\mu\beta^2+\beta^2)r^2 + 2\beta^2 + \mu\beta^4 = 0. \tag{8b}$$

The solutions of Eq. (8b) are

$$r_{1,2}^2 = \frac{1+(1+\mu)\beta^2 \mp \sqrt{1+2(\mu-1)\beta^2 + (1+\mu^2)\beta^4}}{2}. \tag{9}$$

The sum and the product of the solutions of the equations are

$$r_1{}^2 + r_2{}^2 = 1 + \mu\beta^2 + \beta^2,$$

$$r_1{}^2 r_2{}^2 = \frac{2\beta^2 + \mu\beta^4}{2}. \tag{10}$$

Consider Eq. (5) and let $\zeta \to \infty$. The height of the fixed points P and Q can be found as

$$G_{\text{B}} = \sqrt{\frac{r^2}{(1+\mu\beta^2 - r^2)^2}}. \tag{11}$$

Following the procedures used by Den Hartog,[1] the optimum value of β is obtained by $G_{\text{B}}(r_1)=G_{\text{B}}(r_2)$:

$$\frac{r_1}{1+\mu\beta^2 - r_1{}^2} = \frac{-r_2}{1+\mu\beta^2 - r_2{}^2}, \tag{12}$$

leading to

$$\beta_{\text{opt\_B}} = \sqrt{\frac{1-\sqrt{1-2\mu}}{\mu\sqrt{1-2\mu}}}. \tag{13}$$

Substituting r_1 and $\beta_{\text{opt\_B}}$ into Eq. (11) gives the height of the fixed points:

$$G_B(r_1) = G_B(r_2) = G_{B\_\text{opt}} = \sqrt{\frac{\mu\sqrt{1-2\mu}}{1-\mu-\sqrt{1-2\mu}}}. \tag{14}$$

Equation (14) above shows that the velocity amplitude ratio G_{B_opt} of model B with optimum damping and tuning frequency at resonance is zero when the mass ratio is 0.5 while model A can achieve zero velocity amplitude ratio only when the mass ratio approaches infinity.

Brock's[16] approach is applied to find the optimum damping ratio of the absorber in the following. The curve of G_B is required to pass horizontally through the fixed point P which becomes the highest point on the curve of G_B. Consider this curve passing through a point P' of ordinate G_{B_opt} and abscissa $r^2=r_1{}^2+\delta$ with δ approaching zero. Rewriting Eq. (5) as

$$\zeta^2 = \frac{r^2(\beta^2-r^2)^2 - G_B{}^2[(1-r^2)(\beta^2-r^2)-\mu\beta^2r^2]^2}{(2\beta r)^2[G_B{}^2(1+\mu\beta^2-r^2)^2-r^2]} \tag{15}$$

and substituting $r^2=r_1{}^2+\delta$ and $G_B{}^2=\mu\sqrt{1-2\mu}/1-\mu-\sqrt{1-2\mu}$ into the above equation, we would have a result in the form

$$\zeta^2 = \frac{A_0+A_1\delta+A_2\delta^2+\cdots}{B_0+B_1\delta+B_2\delta^2+\cdots}. \tag{16}$$

Since G_B is independent of the damping at the fixed point P, Eq. (16) would assume the indeterminate form 0/0 if $\delta=0$ leading to $A_0=B_0=0$. As δ is a very small number, we may neglect the higher order terms, and the desired result is given by

$$\zeta^2 = \frac{A_1}{B_1}, \tag{17}$$

where

$$A_1 = (r_1{}^2-\beta^2)(3r_1{}^2-\beta^2) - 2G^2((1-r_1{}^2)(\beta^2-r_1{}^2) - \mu\beta^2r_1{}^2)(-1+2r_1{}^2-\beta^2(1+\mu)), \tag{18}$$

$$B_1 = 4\beta^2r_1{}^2(G^2(-2-2\mu\beta^2+2r_1{}^2)-1) + 4G^2(1+\mu\beta^2 - r_1{}^2)^2 - 4r_1{}^2. \tag{19}$$

Now substituting $r_1{}^2$ of Eqs. (5), (8a), (8b), and (9) into Eq. (11), we may write

$$\zeta_{r_1}^2 = \frac{(-2+2\alpha+(10-8\alpha)\mu-7\mu^2)\gamma+4-4\alpha-4\mu+(5\alpha-7)\mu^2+3\mu^3}{\beta^2\mu\alpha(4-4\alpha+4\mu)\gamma+\beta^2\mu\alpha(8-8\alpha-8\mu\alpha-12\mu^2)}, \tag{20}$$

where $\alpha=\sqrt{1-2\mu}$ and $\gamma=\sqrt{2-3\mu^2-(2\mu+2)\sqrt{1-2\mu}}$.

By a similar procedure with $r_2{}^2$ of Eq. (9), we obtain

$$\zeta_{r_2}^2 = \frac{(2-2\alpha-(10-8\alpha)\mu+7\mu^2)\gamma+4-4\alpha-4\mu+(5\alpha-7)\mu^2+3\mu^3}{-\beta^2\mu\alpha(4-4\alpha+4\mu)\gamma+\beta^2\mu\alpha(8-8\alpha-8\mu\alpha-12\mu^2)}. \tag{21}$$

As suggested by Brock,[16] a convenient average value of the damping ratio is selected as the optimum damping ratio which is written as

$$\zeta_{\text{opt\_B}} = \sqrt{\frac{A\sqrt{1-2\mu}+B}{8\mu^2\sqrt{1-2\mu}(-7\mu^2-2\mu+4+(3\mu^2-2\mu-4)\sqrt{1-2\mu})}}, \tag{22}$$

where $A=16-48\mu-12\mu^2+84\mu^3+3\mu^4-15\mu^5$ and $B=-16+64\mu-28\mu^2-112\mu^3+61\mu^4+38\mu^5$.

The velocity amplitude ratios of the primary system for models A and B under optimum tuning and damping with mass ratio $\mu=0.2$ were calculated using Eqs. (1), (2a), (2b), (5), (13), and (22), and the results are plotted in Fig. 3. The maximum non-dimensional velocity amplitude of mass M of model B was 18% smaller than that of model A.

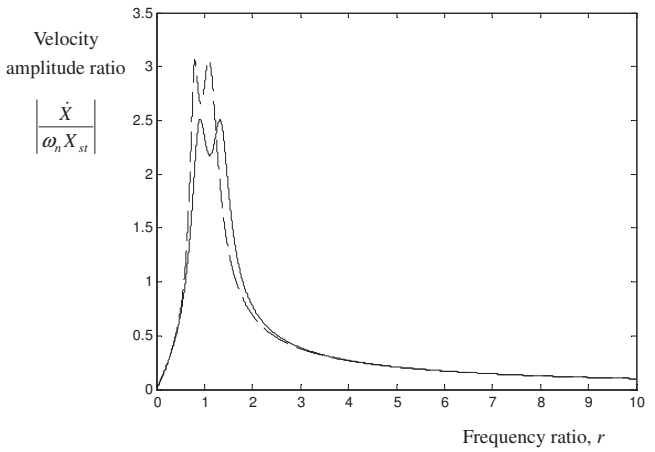


FIG. 3. Comparison of the velocity amplitude ratio between the vibration of the primary system for model A (-----) and model B (solid line) at mass ratio $m/M=0.2$.

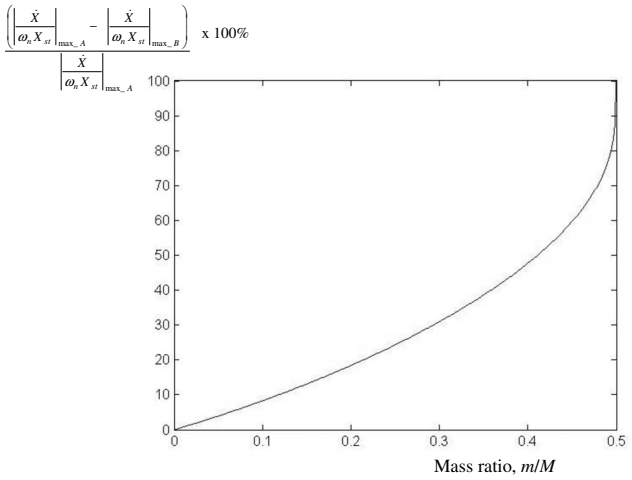


FIG. 4. Percentage reduction in vibration velocity amplitude of M of the proposed absorber relative to the traditional absorber at different mass ratios.

To prove that the proposed absorber is in general superior to the traditional absorber for vibration velocity suppression, the difference of maximum velocity amplitude of the primary system between using the proposed absorber and using the traditional absorber is written as

$$\begin{aligned}
&\left(\left|\frac{\dot{X}_1}{\omega_n X_{\mathrm{st}}}\right|_{\mathrm{max_A}} - \left|\frac{\dot{X}_1}{\omega_n X_{\mathrm{st}}}\right|_{\mathrm{max_B}}\right) \\
&= \sqrt{\frac{2+\mu}{\mu(1+\mu)}} - \sqrt{\frac{\mu\sqrt{1-2\mu}}{1-\mu-\sqrt{1-2\mu}}} \\
&= \left[\frac{2+\mu}{\mu(1+\mu)} - \frac{\mu\sqrt{1-2\mu}}{1-\mu-\sqrt{1-2\mu}}\right] \Bigg/ \left[\sqrt{\frac{2+\mu}{\mu(1+\mu)}} \right. \\
&\quad \left. + \sqrt{\frac{\mu\sqrt{1-2\mu}}{1-\mu-\sqrt{1-2\mu}}}\right] \\
&= \left[\frac{2+\mu}{\mu(1+\mu)} \right. \\
&\quad \left. - \frac{(1-\mu+\sqrt{1-2\mu})\mu\sqrt{1-2\mu}}{(1-\mu)^2-(1-2\mu)}\right] \Bigg/ \left[\sqrt{\frac{2+\mu}{\mu(1+\mu)}} \right. \\
&\quad \left. + \sqrt{\frac{(1-\mu+\sqrt{1-2\mu})\mu\sqrt{1-2\mu}}{(1-\mu)^2-(1-2\mu)}}\right] \\
&= \left[\frac{1-\sqrt{1-2\mu}}{\mu(1+\mu)} + \frac{\mu\sqrt{1-2\mu}}{1+\mu} + 2\right] \Bigg/ \left[\sqrt{\frac{2+\mu}{\mu(1+\mu)}} \right. \\
&\quad \left. + \sqrt{\frac{(1-\mu+\sqrt{1-2\mu})\sqrt{1-2\mu}}{\mu}}\right] > 0, \qquad (23)
\end{aligned}$$

where $0 < \mu \leqslant 0.5$.

The above equation shows that the proposed absorber (model B) provides smaller maximum vibration velocity amplitude of the primary mass M excited by a harmonic force than the traditional absorber (model A) if both absorbers are optimally tuned based on the fixed-point theory. Plotted in Fig. 4 is Eq. (23) which shows the corresponding percentage reduction in velocity amplitude of the mass M of the proposed absorber relative to the traditional absorber at different mass ratios.

IV. CONCLUSION

Optimum tuning condition including the frequency and damping ratios of the proposed absorber has been derived based on the fixed-point theory. Under the optimum tuning condition of the absorbers, it is proved analytically that the proposed absorber provides a larger suppression of maximum vibration velocity response of the primary system than the traditional absorber. The comparison reveals that the maximum velocity amplitude response under the optimized condition of model B is always less than that of model A. The proposed absorber (model B) may be a better alternative absorber design than the traditional damped DVA whenever it is practical to be used. In some applications such as vibration absorption of aircraft panels,[13] the proposed absorber may not be applicable because it is impossible to connect the absorber directly to the ground via a damper. However, the proposed absorber is recommended whenever its configuration is practical for applications requiring the maximum vibration velocity response of the primary system to be minimized.

[1] J. P. Den Hartog, *Mechanical Vibrations* (Dover, New York, 1985).
[2] B. G. Korenev and L. M. Reznikov, *Dynamic Vibration Absorbers, Theory and Technical Applications* (Wiley, New York, 1993).
[3] W. O. Wong, S. L. Tang, Y. L. Cheung, and L. Cheng, "Design of a dynamic vibration absorber for vibration isolation of beams under point or distributed loading," J. Sound Vib. **301**, 898–908 (2007).
[4] Y. L. Cheung and W. O. Wong, "Isolation of bending vibration in a beam structure with a translational vibration absorber and a rotational vibration absorber," J. Vib. Control **14**, 1231–1246 (2008).
[5] M. Z. Ren, "A variant design of the dynamic vibration absorber," J. Sound Vib. **245**, 762–770 (2001).
[6] W. O. Wong and Y. L. Cheung, "Optimal design of a damped dynamic vibration absorber for vibration control of structure excited by ground motion," Eng. Struct. **30**, 282–286 (2008).
[7] D. Karnopp, "Active and semi-active vibration isolation," J. Mech. Des. **117**, 177–185 (1995).
[8] S. K. George and K. Shankar, "Vibrational energies of members in structural networks fitted with tuned vibration absorbers," Int. J. Struct. Stab. Dyn. **6**, 269–284 (2006).
[9] S. J. Esteve and M. E. Johnson, "Adaptive Helmholtz resonators and passive vibration absorbers for cylinder interior noise control," J. Sound Vib. **288**, 1105–1130 (2005).
[10] Z. Yamaguchi, A. Sugimoto, Y. Yano, and J. S. Bolton, "Reduction of radiated sound from a continuum using low-damping vibration absorbers," J. Acoust. Soc. Am. **123**, 3064 (2008).
[11] M. B. Ozer and T. J. Royston, "Passively minimizing structural sound radiation using shunted piezoelectric materials," J. Acoust. Soc. Am. **114**, 1934–1946 (2003).
[12] M. R. Jolly, "Passive tuned vibration absorbers for sound radiation reduction from vibrating panels," J. Sound Vib. **191**, 577–583 (1996).
[13] Z. H. Sun, J. C. Sun, W. Chong, and D. Yang, "Dynamic vibration absorbers used for increasing the noise transmission loss of aircraft panels," Appl. Acoust. **48**, 311–321 (1996).
[14] D. Du, "Analytical solutions for DVA optimization based on the Lyapunov equation," J. Vibr. Acoust. **130**, 054501 (2008).
[15] T. Asami and O. Nishihara, "Closed-form exact solution to H_∞ optimization of dynamic vibration absorbers (application to different transfer functions and damping systems)," J. Vibr. Acoust. **125**, 398–405 (2003).
[16] J. E. Brock, "A note on the damped vibration absorber," J. Appl. Mech. **13**, A284 (1946).

Adaptive plasticity in brainstem of adult listeners following earplug-induced deprivation (L)

Kevin J. Munro[a] and Jennifer Blount
School of Psychological Sciences, University of Manchester, Manchester M13 9PL, United Kingdom

(Received 9 March 2009; revised 8 June 2009; accepted 9 June 2009)

Previous research has shown that loudness perception is modified in adult listeners following 2 weeks of continuous sensory deprivation or stimulation [Formby, C. *et al.* (2003). J. Acoust. Soc. Am. **114**, 55–58]. However, it is not known if the auditory system undergoes physiological changes or if the listeners simply recalibrate their behavioral criteria such that they become more, or less, conservative following sensory deprivation and stimulation, respectively. The results of this study, comparing threshold of the middle ear acoustic reflex in the two ears of adult listeners after use of a unilateral earplug, are consistent with adaptive plasticity. Acoustic reflexes were measured at a lower sound pressure level in the ear that had been plugged for 7 days. Thus, the effect is consistent with a central gain mechanism mediated by a process within the brainstem.
 [DOI: 10.1121/1.3161829]

PACS number(s): 43.64.Ha, 43.64.Ri, 43.64.Bt [BLM] Pages: 568–571

I. INTRODUCTION

Formby *et al.* (2003, 2007) presented evidence that loudness perception is modified in adult listeners following 2–4 weeks of partial auditory deprivation or stimulation. They reported on ten listeners who wore bilateral earplugs (deprivation) or noise generators (stimulation), which provided a shift in hearing thresholds of around 20–30 dB from 2 to 4 kHz. After the earplug experience, listeners required an increase of around 6–8 dB to match pre-treatment loudness measures for moderate and high stimulus presentation levels. Conversely, listeners required a decrease of around 6–8 dB after noise generator experience. The authors explain their findings in terms of a hypothetical gain process that regulates supra-threshold sensitivity in the auditory system. Although the findings were described in terms of adaptive plasticity, it is not known if the auditory system underwent a physiological change. A more prosaic explanation, acknowledged by the authors, is that listeners simply recalibrated their behavioral criteria such that they become more conservative after wearing the earplugs and less conservative after wearing the noise generators.

To differentiate between these two speculative mechanisms, one needs direct evidence of adaptive plasticity in the auditory system. One potential source of evidence would appear to be available from manipulation of the acoustic reflex threshold (ART), which is mediated by a three-nuclei arc within the brainstem. The ascending limb of the acoustic reflex pathway includes the primary auditory neurons and the cochlear nucleus neurons. The mode of transfer of this information to the motor neurons via the facial nuclei involves an intermediate neuron at the level of the medial superior olivary complex (Borg, 1973). There is a known relationship between loudness discomfort and the sound level at which there is a reflex contraction of the muscles in the middle ear (Olsen, 1999). Accordingly, if the ART changes after a period of earplug-induced deprivation, then this finding would provide direct evidence of adaptive plasticity in the mature auditory system.

The aim of this study, therefore, was to investigate ARTs in normal-hearing adult listeners following unilateral earplug experience. The authors hypothesized that the threshold of the acoustic reflex would be the same in both ears before unilateral earplug experience, but they expected the ART to decrease in the ear that underwent a period of earplug-induced deprivation, consistent with adaptive plasticity and an enhanced physiological response to the loudness of the acoustic stimulation.

II. METHODS

A. Participants

Eleven consenting volunteers (mean age 23.6 years; s.d. 2.38) participated in the study, which received ethics approval from the University of Manchester. All participants were screened for normal-hearing sensitivity (<20 dB HL from 0.25 to 8 kHz and no asymmetry >10 dB at any frequency) and normal middle ear function on tympanometry (middle ear pressure +50 to −50 daPa, middle ear compliance 0.3 to 1.5 cm^3).

B. Noise-attenuating earplugs

An aural impression was taken of one ear from each participant (five right and six left) for the purpose of manufacturing a customized noise-attenuating earplug. Earplugs were fabricated from a durable biopore material with a shore rating of 25; a soft material is important for extensive periods of continuous earplug use, as required in this study. Listeners were instructed on use and maintenance of the earplug and were required to demonstrate competent insertion and removal before they were accepted in the study. They were asked to wear the earplug continuously for 7 days, except for daily ablutions.

[a]Author to whom correspondence should be addressed. Electronic mail: kevin.munro@manchester.ac.uk

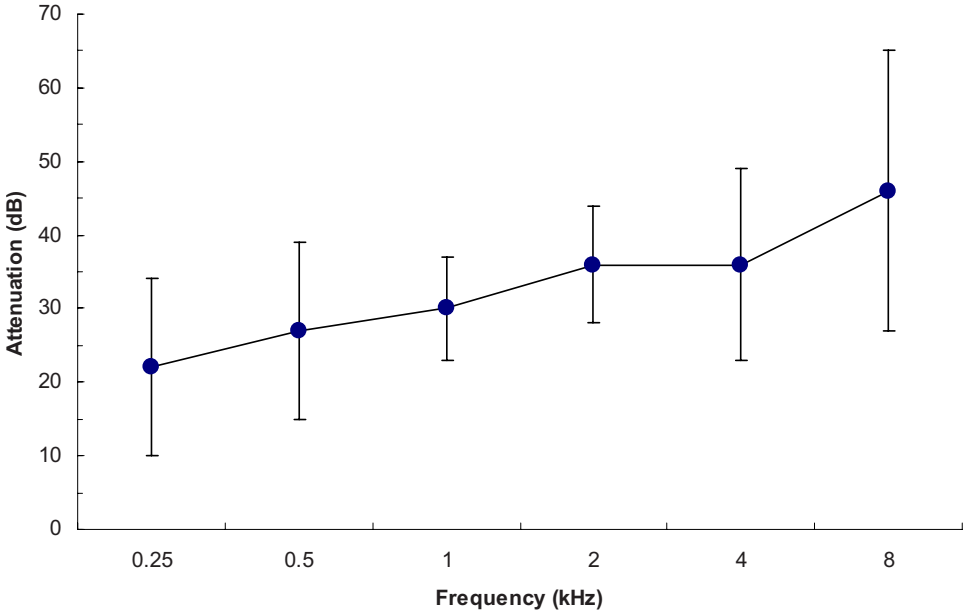


FIG. 1. (Color online) Mean frequency-dependent attenuation provided by the customized earplugs. Error bars show ±1 s.d. (n=11).

Pure-tone hearing thresholds were measured with supra-aural earphones before and after insertion of the earplug. The difference between these measures (occluded minus non-occluded) is shown in Fig. 1. The mean attenuation provided by the earplugs increased from 22 dB at 0.25 kHz to 46 dB at 8 kHz.

C. Acoustic reflex thresholds

ARTs were measured on three occasions over a 14 day period: immediately before use of earplug (day 0), immediately after removal of earplug on day 7 (day 7), and 7 days after removal of earplug (day 14). The ARTs were measured at these same times for the untreated control ears. Ipsilateral ARTs (i.e., eliciting stimuli and reflex measurement in same ear) were measured using a GSI Tympstar middle ear analyzer and 226 Hz probe tone. Measurements were limited to 2 and 4 kHz for which mean earplug attenuation was around 35 dB (the middle ear analyzer used in the study did not provide an 8 kHz stimulus where earplug attenuation was greatest). Stimuli were presented at 70 dB HL and were increased in 5 dB steps until an observable reflex was measured (i.e., a decrease in compliance of $\geq$0.02 cm^3). Reflex growth was confirmed by presenting the stimulus at +5 dB above this level. The level was then reduced by 10 dB and increased in 2 dB step sizes to establish the reflex threshold.

III. RESULTS

The mean ART data (±1 s.d.) are shown in Table I. The same pattern is apparent at both test frequencies. There is little or no asymmetry between ears at 0 and 14 days; however, immediately after 7 days of earplug experience, there are mean asymmetries of +8 at 2 kHz and +10 dB at 4 kHz. At 2 kHz, ten (91%) subjects had higher ARTs in the fitted ear and in seven (64%) the difference was $\geq$8 dB. At 4 kHz, 11 (100%) subjects had higher ARTs in the fitted ear and in 7 (64%) the difference was $\geq$10 dB. The mean difference between ears was analyzed using a two-factor (time [3] and frequency [2]) repeated-measures analysis of variance (ANOVA). The main effect of time was statistically significant [$F(1,10)=24.3$; $p<0.01$]. The main effect of frequency was not statistically significant [$F(1,10)=0.27$; $p=0.62$]. There was no interaction between time and frequency [$F(1.22,12.19)=4.18$; $p=0.07$] indicating that mean ARTs are similar over time for both eliciting stimuli.

Most of the asymmetry immediately after earplug experience (day 7) was due to changes in the plugged ear, but there was also a small change in the control ear. Mean ARTs reduced by 5–7 dB in the plugged ear and increased by 1–3 dB in the control ear. A separate two-factor (time [3] and ear [2]) repeated-measures ANOVA was performed on the mean data for each frequency separately. At 2 kHz, the main effect of ear was statistically significant [$F(1,10)=5.9$; $p=0.03$]. This was due to the asymmetry between ears at the beginning and end of the study. The main effect of time was not statistically significant [$F(1,10)=1.8$; $p=0.19$]. However, there was a statistically significant interaction between ear and frequency [$F(1.22,12.2)=11.42$; $p<0.01$]. The interaction means that the ARTs change differentially for the two ears. A one-factor (time [3]) within-subject ANOVA was carried out on the data for each ear. There was a statistically significant change over time for the plugged ear [$F(2,20)=6.0$; $p<0.01$]. There was also a statistically significant

TABLE I. Mean ARTs (in dB HL) as a function of time. ±1 s.d. is given in parentheses (n=11).

	2 kHz acoustic reflex threshold (dB HL)		
Time (days)	0	7	14
Plugged ear	85 (±8.1)	80 (±8.3)	85 (±7.6)
Control Ear	87 (±8.0)	88 (±8.7)	86 (±7.8)
	4 kHz acoustic reflex threshold (dB HL)		
Time (days)	0	7	14
Plugged ear	88 (±8.5)	81 (±7.4)	89 (±8.4)
Control ear	88 (±8.3)	91 (±8.5)	89 (±7.8)

change over time for the control ear [$F(2,20)=4.4$; $p<0.01$]. Paired sample t-tests were used to compare the data at each time period. There was no statistically significant difference at day 0 [$t(10)=1.02$; $p=0.33$] or day 14 [$t(10)=0.69$; $p=0.51$]; however, there was a statistically significant difference between ears immediately after 7 days of unilateral deprivation [$t(10)=4.2$; $p<0.01$].

At 4 kHz, the main effect of ear and time was not statistically significant [ear: $F(1,10)=4.3$, $p=0.07$; time: $F(1,10)=1.9$, $p=0.17$]. There was a statistically significant interaction between ear and frequency [$F(2,20)=29.2$; $p<0.01$]. A one-factor (time [3]) within-subject ANOVA was then carried out on the data from each ear. There was a statistically significant change over time for the plugged ear [$F(2,20)=11.2$; $p<0.01$]. There was also a statistically significant change over time for the control ear [$F(2,20)=5.1$; $p=0.02$]. Paired sample t-tests were used to compare the data at each time period. There was no statistically significant difference at the beginning [$t(10)=0.21$; $p=0.84$] or at the end of the study [$t(10)=0.10$; $p=0.92$]; however, there was a statistically significant difference between ears after 7 days of unilateral deprivation [$t(10)=6.03$; $p<0.01$].

In summary, both ears show a statistically significant change in mean ART after a 7 day period of unilateral deprivation (ARTs decreased in the ear that had been plugged but increased in the control ear) but return to baseline values within 7 days of earplug removal. There was also a positive correlation between the change in 2 and 4 kHz ARTs at 0 and 7 days ($r=0.83$, $n=11$; $p<0.01$) and there was a negative correlation between the change in 2 and 4 kHz ARTs at 7 and 14 days ($r=0.71$, $n=11$; $p=0.02$).

IV. DISCUSSION

The ART findings reveal direct evidence for adaptive plasticity after a period of unilateral sensory deprivation. This is consistent with the perceptual changes reported by Formby *et al.* (2003, 2007). The earplugs modify the sensory environment sufficiently to induce experience-related changes within the mature auditory system. A study by Decker and Howe (1981) supports this finding. They fitted temporary earplugs to the right ear of 30 adult female listeners and measured ARTs after 10, 20, and 30 h. There was a non-significant trend for the mean ARTs to reduce by around 2 dB at 2 kHz in the plugged ear, irrespective of duration of earplug use.

The changes reported here have been measured at the level of the auditory brainstem. Auditory plasticity in adults has generally been reported at the level of the auditory cortex (Rajan and Irvine, 1998). However, investigators using hearing-impaired adult animals have reported molecular and cellular changes at the level of the brainstem (Illing, 2001; Illing and Reisch, 2006; Willot, 2006) and there is recent evidence of training-induced brainstem plasticity in normal-hearing adult humans. For example, Song *et al.* (2008) used the frequency following response, originating from the auditory midbrain, to demonstrate brainstem plasticity after training native English-speaking adults to incorporate foreign speech sounds into a word identification task. de Boer and Thornton (2008) used contralateral suppression of the click-evoked otoacoustic emission to investigate the involvement of the medial olivo-cochlear (MOC) bundle (which originates in the brainstem and terminates in the cochlea) on perceptual learning. Their results showed that auditory training changes activity of the MOC system. There are at least two lines of evidence of experience-induced changes in the brainstem in listeners with a symmetrical age-related hearing impairment and unilateral hearing aid experience. The first line of evidence is directly related to the present study because ARTs were measured to investigate ear asymmetry (Munro *et al.*, 2007a). The mean ARTs were higher in the hearing aid ear compared to the not-fitted control ear. The lack of baseline data, however, makes it impossible to know whether this asymmetry was due to an increase in the ART measured in the hearing aid ear or a decrease measured in the control ear. In a second study, Munro *et al.* (2007b) investigated ear asymmetry using the auditory brainstem response and showed that the mean peak-to-peak amplitude for wave V to SN_{10} was higher in the hearing aid ear than for the not-fitted control ear. While prospective studies have yet to be reported, the simplest explanation for these findings is that physiological changes occurred as a result of experience with amplified sound.

The findings in the present study provide a physiological basis for the gain control hypothesis discussed by Formby *et al.* (2003, 2007). Because the plugged ear is deprived of input, the central neural processes increase the auditory gain and this increase is revealed by a lower sound level required to elicit an acoustic reflex. This gain mechanism forms the basis for sound therapy desensitization treatment in listeners with hyperacusis, a condition characterized by intolerance to sound levels that are not normally judged as uncomfortable (Jastreboff and Hazell, 1993). It may also explain the ear asymmetry for intensity discrimination and loudness perception in adult listeners with a symmetrical age-related hearing impairment but unilateral hearing aid experience (e.g., Munro and Trotter, 2006; Robinson and Gatehouse, 1995, 1996). Formby *et al.* (2003) reported perceptual changes at frequencies unaffected by the earplug. It is not known if the physiological changes reported in the present study extend to frequencies unaffected by the earplug.

An unexpected finding in this study was the change in the ART in the untreated control ear. The increased ART measured in the control ear reflects an apparent reduction in gain as revealed by a slightly higher sound level required to elicit the acoustic reflex. Darrow *et al.* (2006) showed a complementary binaural effect in mice after unilateral destruction of efferent neurons from the lateral superior olive to the cochlea. The amplitude of the ipsilateral compound action potential from the auditory nerve was enhanced compared to controls, whereas contralateral to the lesion, the amplitude was lower than in controls.

Qiu *et al.* (2000) also provided evidence for increased gain in the deprived auditory system. They destroyed the inner hair cells, the sensory transducers in the cochlea, in adult chinchillas, and showed that activity in the auditory

cortex was unchanged or even enhanced. They interpreted this finding as an increase in central gain to compensate for the reduced input from the cochlea.

V. CONCLUSIONS

This study has shown that partial unilateral sound deprivation, subsequent to prolonged use of earplugs, results in adaptive plasticity in normal-hearing adult listeners. This adaptive plasticity is reversible and can be measured at the level of the auditory brainstem using the ART. The effect is bilateral and complementary, and is consistent with a central gain mechanism. The relationship between these physiological changes and the perceptual consequences await further studies.

ACKNOWLEDGMENT

The authors gratefully acknowledge the assistance of Starkey Laboratories Ltd. who provided the customized earplugs used in the study.

Borg, E. (**1973**). "On the neural organisation of the acoustic middle ear reflex," Brain Res. **49**, 101–123.

Darrow, K. N., Maison, S. F., and Liberman, M. C. (**2006**). "Cochlear efferent feedback balances interaural sensitivity," Nat. Neurosci. **9**, 1474–1476.

de Boer, J., and Thornton, A. R. D. (**2008**). "Neural correlates of perceptual learning in the auditory brainstem: Efferent activity predicts and reflects improvement at a speech-in-noise discrimination task," J. Neurosci. **28**, 4929–4937.

Decker, T. N., and Howe, S. W. (**1981**). "Short-term auditory deprivation: Effect on brainstem electrical response," Hear. Res. **4**, 251–263.

Formby, C., Sherlock, L. P., and Gold, S. L. (**2003**). "Adaptive plasticity of loudness induced by chronic attenuation and enhancement of the acoustic background (L)," J. Acoust. Soc. Am. **114**, 55–58.

Formby, C., Sherlock, L. P., Gold, S. L., and Haweley, M. L. (**2007**). "Adaptive recalibration of central auditory gain," Semin. Hear. **28**, 295–302.

Illing, R.-B. (**2001**). "Activity-dependent plasticity in the adult auditory brainstem," Audiol. Neuro-Otol. **6**, 319–345.

Illing, R.-B., and Reisch, A. (**2006**). "Specific plasticity responses to unilaterally decreased or increased hearing intensity in the adult cochlear nucleus and beyond," Hear. Res. **216–217**, 189–197.

Munro, K. J., Pisareva, N. P., Parker, D. J., and Purdy, S. C. (**2007b**). "Asymmetry in the auditory brainstem response following experience of monaural amplification," NeuroReport **18**, 1871–1874.

Munro, K. J., and Trotter, J. H. (**2006**). "Preliminary evidence of asymmetry in uncomfortable loudness levels after unilateral hearing aid experience: Evidence of functional plasticity in the adult auditory system," Int. J. Audiol. **45**, 684–688.

Munro, K. J., Walker, A. J., and Purdy, S. C. (**2007a**). "Evidence for adaptive plasticity in elderly monaural hearing aid users," NeuroReport **18**, 1237–1240.

Olsen, S. O. (**1999**). "The relationship between the uncomfortable loudness level and the acoustic reflex threshold for pure tones in normally-hearing and impaired listeners-a meta-analysis," Audiology **38**, 61–68.

Jastreboff, P. J., and Hazell, J. W. P. (**1993**). "A neurophysiological approach to tinnitus: Clinical implication," Br. J. Audiol. **27**, 7–17.

Qiu, C.-X., Salvi, R., Ding, D., and Burkhard, R. (**2000**). "Inner hair cell loss leads to enhanced response amplitudes in auditory cortex of unanesthetized chinchillas: Evidence for increased system gain," Hear. Res. **139**, 153–171.

Rajan, R., and Irvine, D. R. F. (**1998**). "Absence of plasticity of the frequency map in dorsal cochlear nucleus of adult cats after unilateral partial cochlear lesion," J. Comp. Neurol. **399**, 35–46.

Robinson, K., and Gatehouse, S. (**1995**). "Changes in intensity discrimination following monaural long-term use of a hearing aid," J. Acoust. Soc. Am. **97**, 1183–1190.

Robinson, K., and Gatehouse, S. (**1996**). "The time course of effects on intensity discrimination following monaural fitting of hearing aids," J. Acoust. Soc. Am. **99**, 1255–1258.

Song, J. H., Skoe, E., Wong, P. C. M., and Kraus, N. (**2008**). "Plasticity in the adult human auditory brainstem following short-term linguistic training," J. Cogn Neurosci. **20**, 1892–1902.

Willott, J. F. (**2006**). "Central auditory plasticity in mouse models of progressive sensorineural hearing loss," *Programming the Central Cortex*, edited by S. Lomber and J. Eggermont (Oxford University Press, Oxford), pp. 181–192.

Nonlinear parabolic equation model for finite-amplitude sound propagation over porous ground layers

Thomas Leissing,[a] Philippe Jean, and Jérôme Defrance
Centre Scientifique et Technique du Bâtiment (CSTB), Université Paris-Est, 24 rue Joseph Fourier, 38400 Saint-Martin-d'Hères, France

Christian Soize
Laboratoire Modélisation et Simulation Multi Echelle, Université Paris-Est, MSME FRE3160 CNRS, 5 bd Descartes, 77454 Marne-la-Vallee, France

(Received 17 December 2008; revised 28 May 2009; accepted 3 June 2009)

The nonlinear parabolic equation (NPE) is a time-domain method widely used in underwater sound propagation applications. It allows simulation of weakly nonlinear sound propagation within an inhomogeneous medium. So that this method can be used for outdoor sound propagation applications it must account for the effects of an absorbing ground surface. The NPE being formulated in the time domain, complex impedances cannot be used and, hence, the ground layer is included in the computational system with the help of a second NPE based on the Zwikker–Kosten model. A two-way coupling between these two layers (air and ground) is required for the whole system to behave correctly. Coupling equations are derived from linearized Euler's equations. In the frame of a parabolic model, this two-way coupling only involves spatial derivatives, making its numerical implementation straightforward. Several propagation examples, both linear or nonlinear, are then presented. The method is shown to give satisfactory results for a wide range of ground characteristics. Finally, the problem of including Forchheimer's nonlinearities in the two-way coupling is addressed and an approximate solution is proposed.
 [DOI: 10.1121/1.3158937]

PACS number(s): 43.25.Cb, 43.25.Jh, 43.28.En, 43.28.Js [RR] Pages: 572–581

I. INTRODUCTION

Simulating outdoor shock wave propagation requires numerical models, which in addition to high-amplitude effects can account for meteorological conditions and ground properties. Indeed meteorological and ground effects can largely affect sound propagation and hence the wave temporal signature. These comprehensive propagation models are generally associated with an intense numerical effort. In this work a nonlinear parabolic equation (NPE) model is used to simulate finite-amplitude sound propagation. The NPE has first been developed by McDonald and Kuperman[1] in 1987 and has been successfully used for underwater acoustics simulations.[2] It has also been coupled to other numerical methods to simulate blast wave propagation in air.[3–6] The NPE is based on the resolution of a nonlinear wave equation over a moving-window that surrounds the wavefront. While reducing domain size, and thus computational cost, the moving-window principle prevents backward propagation. For the derivation of the original NPE model, the reader may refer to articles by McDonald and co-workers[7,8] or Caine and West.[9] The NPE model for a two-dimensional (2D) domain with Cartesian coordinates (x,z) is based on the following equation:

$$D_t R = -\partial_x\left(c_1 R + c_0\frac{\beta}{2}R^2\right) - \frac{c_0}{2}\int \partial_z^2 R dx, \tag{1}$$

where ∂_i means partial derivation with respect to variable i, x is the main propagation direction, z is the transverse propagation direction, and t is the time variable. The ambient sound speed is c_0 while c_1 is the sound speed perturbation in the window, i.e., $c_1 = c(x,z) - c_0$, where $c(x,z)$ is the spatially-dependent sound speed. $R = \rho'/\rho_0$ is a dimensionless over-density variable, with ρ' representing the acoustic density perturbation and ρ_0 the ambient medium density. For air, the coefficient of nonlinearity β is calculated with the help of the ratio of specific heats γ, i.e., $\beta = (\gamma+1)/2$. The first term on the right hand side of Eq. (1) represents refraction and nonlinear effects; the second term accounts for propagation in the transverse direction. D_t is a moving-window operator and is defined by

$$D_t = \partial_t + c_0\partial_x. \tag{2}$$

Note that in Eq. (1), the azimuthal spreading term $c_0 R/(2r)$ has been dropped from the original NPE.[1] The assumptions used to derive this model are (i) weak nonlinearities; (ii) weak sound speed perturbations, i.e., $c_1 \ll c_0$; and (iii) propagation along a main direction. Equation (1) can be used to propagate weak shocks over moderate distances within a domain with spatially-varying sound speed. Various modifications and additions to this original model have been made during the past two decades: Spherical and cylindrical coordinate system versions[10] and high-angle formulation[11] have

[a]Author to whom correspondence should be addressed. Electronic mail: thomas.leissing@cstb.fr

been developed, and Too and Lee[12] extended the NPE with an additional term to account for thermoviscous effects. Propagation in multiple media[13] and propagation through atmospheric turbulences[14] were also successfully studied using this model.

Euler's equations' methods can provide complete solutions to nonlinear sound propagation problems:[15] Realistic absorption models,[16] meteorological effects,[17,18] hilly terrain,[19] and ground impedances[20–23] can be accounted for in a very accurate way. Moreover, it does not suffer from the parabolic approximation inherent to NPE models. However, for long-range wave propagation problems, computational times will often be on the order of days for three-dimensional domains. Despite increasing computational resources and the existence of modern numerical techniques such as the use of efficient absorbing layers[24] or adaptive mesh refinement methods,[25] Euler's equations-based models cannot compete in calculation time with NPE-based methods. Indeed the use of a single variable wave equation makes the NPE an efficient tool for long-range sound propagation simulations. The main motivation for the development of the NPE model presented here is its use to study finite-amplitude wave propagation over urban environments. Reduced computational times will allow this model to be used several hundred times to obtain statistical information on the wave fields.

In the present work, a parabolic equation model, which takes into account the effects of a soft ground layer on sound propagation, is proposed. It is not the objective here to study the propagation *within* the ground layer but rather to capture the effects of the non-rigid interface on the air acoustic fields. The paper is organized as follows. The derivation of the NPE model for porous ground layers is described in Sec. II. The combination of two-way coupling equations presented in Sec. III and two NPE models for atmospheric and porous ground media allows simulation of finite-amplitude sound propagation over an impedant ground surface. Several propagation examples are then shown in Sec. IV and, finally, an approximate solution to include Forchheimer's nonlinearities[26,27] in the two-way coupling is presented in Sec. V.

II. NPE MODEL FOR RIGIDLY-FRAMED POROUS MEDIA

The domain considered is 2D with main axes x (horizontal direction) and z (vertical direction). Total density ρ_T and total pressure p_T variables are noted as follows:

$$\rho_T = \rho_0 + \rho', \quad p_T = p_0 + p', \tag{3}$$

where ρ_0 and p_0 are ambient air density and ambient air pressure, respectively, and ρ' and p' are acoustic perturbations of these quantities. Components of the flow velocity vector $\mathbf{V}$ in the x- and z-directions are u and w, respectively.

In order to preserve one of the most interesting features of NPE models, namely, a short simulation time, including the porous medium into the computational system, must not dramatically increase computational times. It is thus proposed to derive a parabolic equation model similar to Eq. (1), which uses a minimal parametrization: The layer is assumed to be equivalent to a continuous fluid medium. A wave causes a vibration of air particles contained in the ground pores, while the ground frame does not vibrate. The NPE model for sound propagation in porous ground media is based on a nonlinear extension of the Zwikker–Kosten (ZK) model,[28] characterized by a set of four parameters: the dc flow resistivity σ_0, the porosity Ω_0, the tortuosity Φ, which is defined as the ratio of a curved path length to the distance between its end points, and the Forchheimer's nonlinearity parameter ξ. These quantities are assumed to be fixed in space and time. Considering these assumptions equations of continuity and conservation of momentum are[29–31]

$$\partial_t \rho_T + \partial_x(\rho_T u) + \partial_z(\rho_T w) = 0, \tag{4a}$$

$$\Phi\partial_t(\rho_T u) + \partial_x(p_T + \Phi\rho_T u^2) + \partial_z(\Phi\rho_T uw) + \sigma_0\Omega_0(1 + \xi|u|)u = 0, \tag{4b}$$

$$\Phi\partial_t(\rho_T w) + \partial_z(p_T + \Phi\rho_T w^2) + \partial_x(\Phi\rho_T uw) + \sigma_0\Omega_0(1 + \xi|w|)w = 0. \tag{4c}$$

As one can see in Eqs. (4a)–(4c), the tortuosity Φ reduces the pressure gradients and flow resistive terms' amplitude. Combining Eqs. (4a)–(4c) gives

$$\begin{aligned}\Phi\partial_t^2\rho_T = {} & \partial_x^2(p_T + \Phi\rho_T u^2) + \partial_z^2(p_T + \Phi\rho_T w^2) \\ & + 2\partial_x\partial_z(\Phi\rho_T uw) + \sigma_0\Omega_0\partial_x[(1 + \xi|u|)u] \\ & + \sigma_0\Omega_0\partial_z[(1 + \xi|w|)w].\end{aligned} \tag{5}$$

Since the propagation is mainly along the x-axis, only linear terms in z-derivatives are kept in Eq. (5): Terms $\partial_x\partial_z(\Phi\rho_T uw)$, $\partial_z^2(\Phi\rho_T w^2)$, and $\sigma_0\Omega_0\partial_z(\xi|w|w)$ are neglected. Moreover, only terms of order up to two in x-derivatives are retained: The quantity $\partial_x^2(\Phi\rho' u^2)$ is discarded; this leads to

$$\begin{aligned}\Phi\partial_t^2\rho_T = {} & \partial_x^2(p_T + \Phi\rho_0 u^2) + \partial_z^2 p_T + \sigma_0\Omega_0\partial_x[(1 + \xi|u|)u] \\ & + \sigma_0\Omega_0\partial_z w.\end{aligned} \tag{6}$$

To find an expression for the flow velocities u and w the authors use the perturbation expansion method. The same scalings and expansions as in Refs. 1 and 7 are used (however, note that the window speed in the ground layer is set to $c_0/\sqrt{\Phi}$) as follows:

$$x \to x - \frac{c_0}{\sqrt{\Phi}}t, \quad z \to \epsilon^{1/2}z, \quad t \to \epsilon t. \tag{7}$$

The scaling of z by a factor of $\epsilon^{1/2}$ emphasizes the predominance of the propagation in the x-direction. The partial derivatives associated with Eq. (7) are

$$\partial_x \to \partial_x, \quad \partial_z \to \epsilon^{1/2}\partial_z, \quad \partial_t \to \epsilon\partial_t - \frac{c_0}{\sqrt{\Phi}}\partial_x. \tag{8}$$

The dependent variables are expanded as follows:

$$\rho_T \to \rho_0 + \epsilon\rho_1 + \epsilon^2\rho_2 + \epsilon^3\rho_3 + \cdots, \tag{9a}$$

$$u \to \epsilon u_1 + \epsilon^{3/2}u_2 + \epsilon^2 u_3 + \cdots, \tag{9b}$$

$$w \to \epsilon w_1 + \epsilon^{3/2} w_2 + \epsilon^2 w_3 + \cdots . \tag{9c}$$

Inserting Eqs. (8) and (9a)–(9c) in Eq. (4a) gives

$$\left(\epsilon\partial_t - \frac{c_0}{\sqrt{\Phi}}\partial_x\right)(\rho_0 + \epsilon\rho_1' + \epsilon^2\rho_2' + \cdots) = -\partial_x[(\rho_0 + \epsilon\rho_1 + \epsilon^2\rho_2 + \cdots)(\epsilon u_1 + \epsilon^{3/2}u_2 + \cdots)] - \epsilon^{1/2}\partial_z[(\rho_0 + \epsilon\rho_1 + \epsilon^2\rho_2 + \cdots)(\epsilon w_1 + \epsilon^{3/2}w_2 + \cdots)]. \tag{10}$$

A first-order approximation of the flow velocity components can be obtained by equating terms of orders ϵ and $\epsilon^{3/2}$ in Eq. (10) as follows:

$$u_1 = \frac{c_0}{\sqrt{\Phi}}\frac{\rho_1}{\rho_0}, \quad w_1 = 0. \tag{11}$$

Note that $\rho' = \rho_1 + O(\epsilon^2)$, $u = u_1 + O(\epsilon^{3/2})$, and $w = w_1 + O(\epsilon^{3/2})$. The substitution of u and w by u_1 and w_1 in Eq. (6) leads to an error consistent with the accuracy sought. One obtains

$$\Phi\partial_t^2\rho_T = \partial_x^2\left(p_T + c_0^2\frac{\rho'^2}{\rho_0}\right) + \partial_z^2 p_T + \frac{\sigma_0\Omega_0 c_0}{\rho_0\sqrt{\Phi}}\partial_x\left[\left(1 + \frac{\xi c_0}{\sqrt{\Phi}}\left|\frac{\rho'}{\rho_0}\right|\right)\rho'\right]. \tag{12}$$

In order to reduce Eq. (12) to a single variable equation, the total pressure p_T is substituted by a second-order expansion in ρ' from an assumed adiabatic equation of state

$$p_T = p_0 + c_0^2\rho' + c_0^2\left(\frac{\gamma - 1}{2\rho_0}\right)\rho'^2, \tag{13}$$

where γ is the ratio of specific heats. Inserting Eq. (13) in Eq. (12) yields

$$\Phi\partial_t^2\rho' = c_0^2\partial_x^2\left[\rho' + \left(\frac{\gamma+1}{2\rho_0}\right)\rho'^2\right] + c_0^2\partial_z^2\rho' + \frac{\sigma_0\Omega_0 c_0}{\rho_0\sqrt{\Phi}}\partial_x\left[\left(1 + \frac{\xi c_0}{\sqrt{\Phi}}\left|\frac{\rho'}{\rho_0}\right|\right)\rho'\right]. \tag{14}$$

The one-way propagation hypothesis is introduced with a moving-frame operator $D_t^\star$ as follows:

$$D_t^\star = \partial_t + \frac{c_0}{\sqrt{\Phi}}\partial_x. \tag{15}$$

The first-order parabolic approximation gives[9]

$$\partial_t^2 \to -2\frac{c_0}{\sqrt{\Phi}}D_t^\star\partial_x + \frac{c_0^2}{\Phi}\partial_x^2. \tag{16}$$

Replacing the second time derivative in Eq. (14) gives a NPE model for propagation in a porous medium

$$D_t^\star R = -\frac{c_0}{\sqrt{\Phi}}\partial_x\left(\frac{\beta}{2}R^2\right) - \frac{c_0}{2\sqrt{\Phi}}\int\partial_z^2 R dx - \frac{\sigma_0\Omega_0}{2\Phi\rho_0}\left(1 + \frac{\xi c_0}{\sqrt{\Phi}}|R|\right)R. \tag{17}$$

Equation (17) can be used to simulate sound propagation within a porous ground layer. However, if one wants to couple air/ground models, a last modification must be done. Indeed, both models use different moving-window speeds: c_0 and $c_0/\sqrt{\Phi}$. Correcting for the frame-speed difference leads to the following substitution:

$$D_t^\star \to D_t + \frac{c_0}{\sqrt{\Phi}}(1 - \sqrt{\Phi})\partial_x. \tag{18}$$

Equation (17) becomes

$$D_t R = -\frac{c_0}{\sqrt{\Phi}}\partial_x\left[(1 - \sqrt{\Phi})R + \frac{\beta}{2}R^2\right] - \frac{c_0}{2\sqrt{\Phi}}\int\partial_z^2 R dx - \frac{\sigma_0\Omega_0}{2\Phi\rho_0}\left(1 + \frac{\xi c_0}{\sqrt{\Phi}}|R|\right)R. \tag{19}$$

The NPE model described by Eq. (19) is able to simulate finite-amplitude sound propagation *within* a rigidly-framed porous material described by a set of four parameters. Note that if one sets $\Phi = 1$ and neglects losses in the layer, i.e., $\sigma_0 = 0$, the model exactly reduces to the usual NPE model for atmospheric propagation given in Eq. (1). Equation (19) allows to draw some conclusions about finite-amplitude sound propagation in porous media: (i) The sound speed in the medium is inversely proportional to the square root of the material tortuosity, i.e., $c = c_0/\sqrt{\Phi}$; (ii) the attenuation in the ground layer is composed of a linear term plus a nonlinear term; and (iii) with the hypothesis used, the material resistivity is proportional to the overdensity R.

III. DERIVATION OF TWO-WAY COUPLING EQUATIONS

As NPE models for air and ground layers use the same moving-frame speed, they can be combined to simulate finite-amplitude sound propagation over a rigidly-framed porous ground layer. This section aims at establishing first-order coupling equations to link these two parabolic propagation models. In the following the authors assume that the deformation of the interface by the wave is small.[13]

A. Derivation

An air layer, whose fields are noted p'^a, u^a, and w^a, is considered. To construct the air-ground interfacial condition a rigidly-framed porous ground layer is introduced; its fields are noted p'^g, u^g, and w^g. With these notations interfacial boundary conditions are continuity of pressure and normal flow velocity

$$[p'^a] = [p'^g], \quad [w^a] = [w^g], \tag{20}$$

where the square brackets denote the field quantity on the air-ground interface. Expressions of w^a and w^g involving the pressure disturbance p' *to the first order* are sought; linearized equations are hence used. For the air layer the authors use the linearized Euler's equation

$$\rho_0\partial_t(w^a) = -\partial_z p_T^a. \tag{21}$$

The perturbation expansion method is used and the same scalings as in Sec. II and in Refs. 1 and 7 are used. Rewriting Eq. (21) and equating terms of orders 1 and 3/2 give

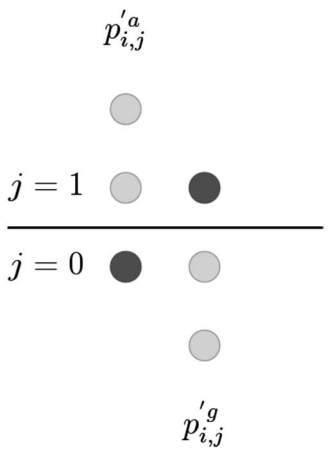


FIG. 1. The air-ground interface is taken to be midway between two vertical grid points with indices $j=0$ and $j=1$. Auxiliary virtual points (black circles) with pressures $p'^a_{i,0}$ and $p'^g_{i,1}$ are created.

$$w^a_1 = 0, \quad w^a_2 = (\rho_0 c_0 \partial_x)^{-1} \partial_z p'^a_1. \tag{22}$$

Note that $w^a = w^a_1 + w^a_2 + O(\epsilon^{5/2})$. To the order of accuracy sought in this work it can be written as

$$w^a = (\rho_0 c_0 \partial_x)^{-1} \partial_z p'^a_1. \tag{23}$$

To find an expression for w^g the authors start from the following momentum equation:[20]

$$\Phi \rho_0 \partial_t w^g = -\Omega_0 \partial_z p^g_T - \sigma_0 \Omega_0 w^g. \tag{24}$$

The same procedure is applied; one finds

$$w^g = (\sqrt{\Phi}\rho_0 c_0 \partial_x - \sigma_0 \Omega_0)^{-1} \Omega_0 \partial_z p'^g_1. \tag{25}$$

The interfacial condition for the continuity of vertical velocities w^a and w^g can now be written as

$$[(\rho_0 c_0 \partial_x)^{-1} \partial_z p'^a] = [(\sqrt{\Phi}\rho_0 c_0 \partial_x - \sigma_0 \Omega_0)^{-1} \Omega_0 \partial_z p'^g]. \tag{26}$$

Rearranging Eq. (26) leads to

$$\left[\sqrt{\Phi}\partial_z p'^a - \frac{\sigma_0 \Omega_0}{\rho_0 c_0} \int \partial_z p'^a dx\right] = [\Omega_0 \partial_z p'^g]. \tag{27}$$

B. Discretization

The variables $p'^a_{i,j}$ and $p'^g_{i,j}$ are introduced to denote pressures in layer a (air layer) and layer g (porous ground layer), respectively, at range $i\Delta x$ *in the moving-window* and altitude $j\Delta z$. The air-ground interface is taken to be midway between two vertical grid points with indices $j=0$ and $j=1$. Auxiliary virtual points with pressures $p'^a_{i,0}$ and $p'^g_{i,1}$ are created. Figure 1 shows a sketch of the configuration.

A trapezoidal law and finite-difference expressions for p'^a and p'^g and their derivatives are used to discretize Eq. (27). For a generic layer l the authors use

$$[p'^l] = \frac{p'^l_{i,1} + p'^l_{i,0}}{2}, \tag{28a}$$

$$[\partial_z p'^l] = (p'^l_{i,1} - p'^l_{i,0})\Delta z^{-1}. \tag{28b}$$

Replacing these approximations with Eq. (27) and using the condition of pressure equality across the interface finally give expressions for unknown quantities $p'^a_{i,0}$ and $p'^g_{i,1}$ as follows:

$$(A+G)p'^a_{i,0} = (A-G)p'^a_{i,1} + 2Gp'^g_{i,0} + S\sum_{m=N_x}^{i+1}(p'^a_{m,1} - p'^a_{m,0}), \tag{29a}$$

$$(A+G)p'^g_{i,1} = (G-A)p'^g_{i,0} + 2Ap'^a_{i,1} + S\sum_{m=N_x}^{i+1}(p'^a_{m,1} - p'^a_{m,0}), \tag{29b}$$

where N_x is the number of points in the moving-window in the x-direction and

$$A = \sqrt{\Phi} + \tfrac{1}{2}S, \tag{30a}$$

$$G = \Omega_0, \tag{30b}$$

$$S = \frac{\sigma_0 \Omega_0 \Delta x}{c_0 \rho_0}. \tag{30c}$$

Equations (29) and (30) give expressions for the unknown pressures $p^a_{i,0}$ and $p^g_{i,1}$ and thus allow, used together with the atmospheric and porous ground NPE models, to simulate weakly nonlinear sound propagation over an impedant ground.

C. Properties

In this section, fundamental properties of the boundary conditions are described and some notes about its numerical implementation are given.

Limitations. First-order formulations of the constitutive equations have been used to derive the boundary interface condition. This implies that nonlinearities cannot be taken into account in the two-way coupling.

Causality. The x-integral present in NPE models [see, for example, Eq. (1)] is calculated from right to left in the calculation grid, and the same method is used for coupling [note the reversed sum indices in Eqs. (29a) and (29b)]. This ensures that no perturbation is introduced ahead of the point where the wave hits the ground, and thus implies that the interfacial condition is causal.

Consistency with classical boundary conditions. If one sets $\Phi=+\infty$ one obtains the following from Eqs. (29a) and (29b): $p'^a_{i,0}=p'^a_{i,1}$, which is, with the discretization used, the condition for a totally rigid interface. A transparent interface condition can be obtained by setting $\sigma_0=0$, $\Omega_0=1$, and $\Phi=1$ (parameters for an air layer). This leads to $A=1$ and $G=1$ and thus $p'^a_{i,0}=p'^g_{i,0}$ and $p'^g_{i,1}=p'^a_{i,1}$, which is the condition for perfect transmission. If one sets $\sigma=0$ and $\Omega_0=1$, Eqs. (29a) and (29b) become

$$p'^a_{i,0} = \frac{\sqrt{\Phi}-1}{\sqrt{\Phi}+1}p'^a_{i,1} + \frac{2}{\sqrt{\Phi}+1}p'^g_{i,0}, \tag{31a}$$

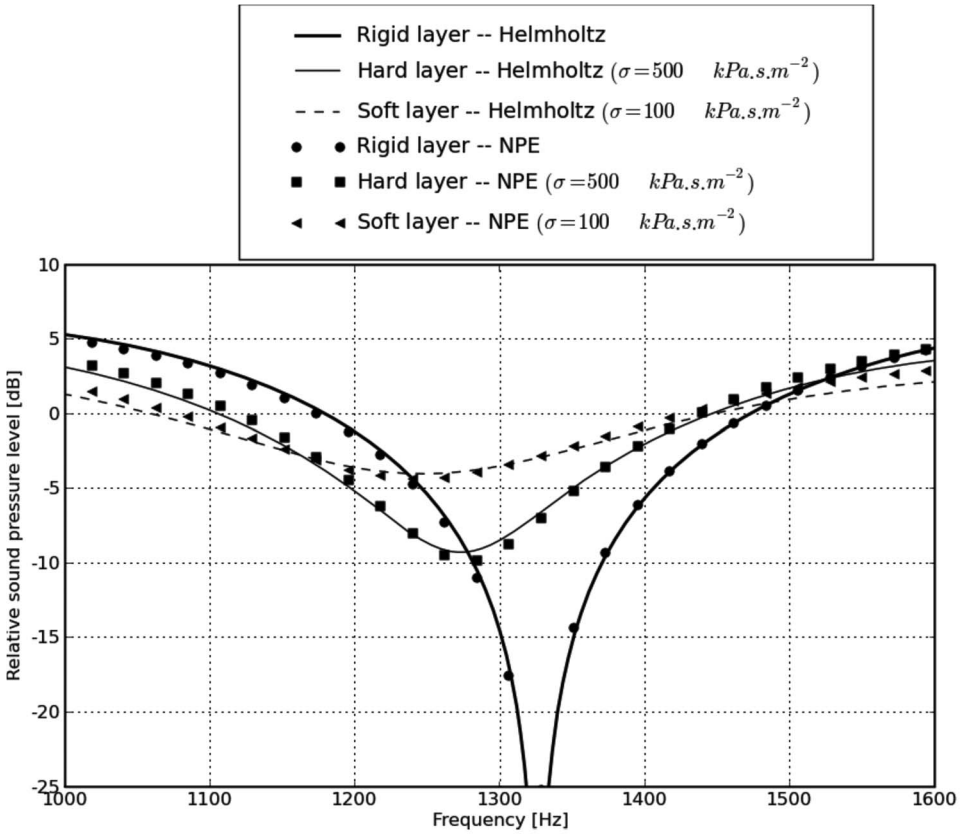


FIG. 2. SPLs relative to free field at the receiver for the three different ground layers, for NPE and analytical solutions. The source and receiver are placed at $(x_s,z_s)=(0,1.4)$ m and $(x_r,z_r)=(10,1.4)$ m.

$$p'^{g}_{i,1}=\frac{1-\sqrt{\Phi}}{\sqrt{\Phi}+1}p'^{g}_{i,0}+\frac{2\sqrt{\Phi}}{\sqrt{\Phi}+1}p'^{a}_{i,1}, \tag{31b}$$

which is the interface condition for two fluid layers with densities ρ_0 and $\sqrt{\Phi}\rho_0$.[13]

Numerical implementation. To solve the diffraction operator a first-order finite-difference approximation for spatial discretization and the Crank–Nicolson method for time marching are used. This leads to a tridiagonal system of equations that is solved columnwise, from right to left in the calculation grid. The boundary interface condition can thus be naturally included in the diffraction solver by imposing values on corresponding points without any additional solver modifications.

IV. NUMERICAL EXAMPLES

In this section, numerical examples of sound propagation over porous ground layers are presented to illustrate the coupling method and to evaluate its performances.

A. Linear propagation

1. Reference solutions

The solutions of the 2D Helmholtz equation are used as references. In the case of 2D wave propagation over a flat and impedant ground surface the pressure is given by

$$p_r=i\pi H_0^{(1)}(kR_1)+Qi\pi H_0^{(1)}(kR_2), \tag{32}$$

where p_r is the complex pressure at the receiver, k is the wavenumber, R_1 and R_2 are the source-receiver and image source-receiver distances, respectively, and $H_0^{(1)}$ is the Hankel function of the first kind and of order zero. Q is the cylindrical reflection coefficient; it can be calculated with the help of Laplace transforms.[20,32] The normalized impedance used to calculate the reflection coefficient is[20]

$$Z=\sqrt{\frac{\Phi}{\Omega_0^2}+i\frac{\sigma_0}{\rho_0\Omega_0\omega}}. \tag{33}$$

2. Configuration

The sound speed is constant within the domain ($c_0=340$ m s^{-1}) and absorption from air is not included in the model. The source is positioned at $(x_s,z_s)=(0.0,1.4)$ m. The signal used is a sine pulse with wavelength $\lambda=0.27$ m ($f=1259.25$ Hz) and its peak amplitude is low enough for the propagation to be considered linear. A receiver is placed 10 m away from the source at $(x_r,z_r)=(10,1.4)$ m. The source and receiver positions ensure that the authors are within the parabolic equation angular validity domain (the angle from source to image-receiver is $\theta\approx 15°$). Spatial steps are equal to 7.5×10^{-3} m in both directions, thus giving a spatial resolution of about 36 points/λ, ensuring sufficient resolution at higher frequencies and near the air/ground boundary. The time step is $\Delta x/c_0$, so that at each time step the window advances one spatial step. Since the Crank–Nicolson method is used the numerical scheme is stable. Three different ground layers of thickness 1 m are considered. The first ground layer is a perfectly rigid surface ($\Phi\gg1$). The second and third layers have identical tortuosity ($\Phi=3$) and porosity ($\Omega_0=0.3$), but different flow resistivities (σ_0

$=500$ kPa s m^{-2} and $\sigma_0=100$ kPa s m^{-2}). The NPE window including the ground layer is 3 m wide and 4.125 m high (400×550 points).

3. Results

Two modifications are done on signals at the receiver: First, in order to obtain a free field reference, time histories are cropped after the direct wave. Next the trailing part of time signals is cropped after the reflected wave to suppress the low-amplitude numerical oscillations. Let us denote the complex pressures at the receiver and in the free field by p_r and p_{free}, respectively. The relative sound pressure level (SPL) ΔL is then calculated with

$$\Delta L = 10\log\left(\frac{p_r^2}{p_{\text{free}}^2}\right). \tag{34}$$

Note that Salomons *et al.*[20] found very little difference on relative SPLs when comparing sources with different decay rates. Analytical solutions for cylindrical line sources are thus valid references for comparison with the NPE model used in this work.

Relative SPLs at the receiver are shown in Fig. 2, for both analytical and NPE calculations. Very good agreement can be observed, independently of the ground properties: Even for the softest layer ($\sigma_0=100$ kPa s m^{-2}) the difference between analytical and NPE calculations is at most 1 dB. The frequencies where negative interference occurs are 1325, 1273, and 1246 Hz for the rigid case, the ground layer with $\sigma_0=500$ kPa s m^{-2}, and the ground layer with $\sigma_0=100$ kPa s m^{-2}, respectively. As one can see in Fig. 2 the NPE model presented does not only accurately recreate reflected wave amplitude decrease, but does account for the change in least reflective frequencies due to the additional delay given during reflection.

B. Nonlinear propagation

1. Reference solution

To obtain reference results to compare to the NPE model simulations in the case of high-amplitude waves, solutions of Euler's equations are used. The computational domain is composed of an air layer and a ground layer. In a 2D Cartesian coordinate system the constitutive equations for the air layer are

$$\partial_t\rho_T + \partial_x(\rho_T u) + \partial_z(\rho_T w) = 0, \tag{35a}$$

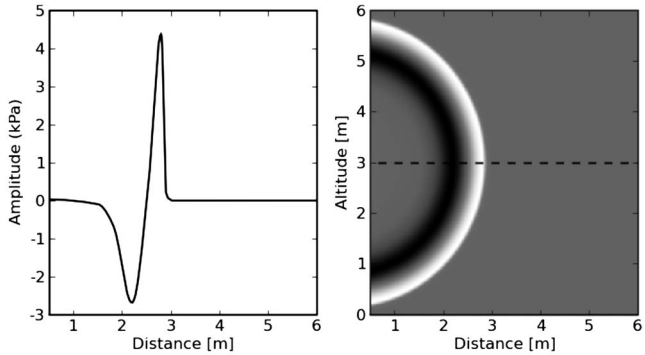


FIG. 3. Left: pressure waveform at altitude $z=3$ m. Right: initial pressure waveform used to start the reference and NPE calculations.

$$\partial_t(\rho_T u) + \partial_x(\rho_T u^2) + \partial_z(\rho_T uw) = -\partial_x p_T, \tag{35b}$$

$$\partial_t(\rho_T w) + \partial_x(\rho_T uw) + \partial_z(\rho_T w^2) = -\partial_z p_T, \tag{35c}$$

$$\partial_t(\rho_T e_0) + \partial_x(\rho_T u e_0) + \partial_z(\rho_T w e_0) = -\partial_x(p_T u) - \partial_z(p_T w), \tag{35d}$$

where e_0 is the energy per unit mass. Within the ground layer momentum conservation equations write

$$\Phi\partial_t(\rho_T u) + \partial_x(p_T + \Phi\rho_T u^2) + \partial_z(\Phi\rho_T uw) + \sigma_0\Omega_0(1 + \xi|u|)u = 0, \tag{36a}$$

$$\Phi\partial_t(\rho_T w) + \partial_z(p_T + \Phi\rho_T w^2) + \partial_x(\Phi\rho_T uw) + \sigma_0\Omega_0(1 + \xi|w|)w = 0. \tag{36b}$$

The energy equation (37) and the ideal gas law (38) close the equation system

$$\rho_T e_0 = \rho_T C_v T + \frac{\rho_T|\mathbf{V}|^2}{2}, \tag{37}$$

$$p_T = \rho_T RT, \tag{38}$$

where T is the gas temperature, C_v is the specific heat capacity at constant volume, and R is the gas constant. To solve this equation system a weighted essentially non-oscillatory (WENO) algorithm[33] for space discretization and a third-order total variation diminishing scheme[34] for time marching are used. These numerical algorithms are briefly presented in the Appendix.

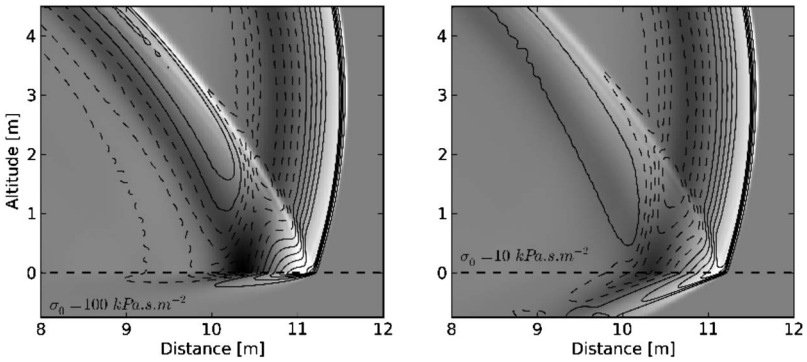


FIG. 4. Snapshots at time $t=33$ ms. Left: $\sigma_0=100$ kPa s m^{-2}. Right: $\sigma_0=10$ kPa s m^{-2}. Colormap: solution from NPE model. Contour lines: solution from Euler's equations. Ten contour lines equally spaced from -800 to 800 Pa are shown. Contours corresponding to negative values are represented by dashed lines, and positive ones by solid lines.

2. Configuration

In this example, standard atmospheric conditions are used ($T=293$ K, $\rho_0=1.2$ kg m^{-3}, and $p_0=1.03\times10^5$ Pa). The source is positioned at $(x_s,z_s)=(0,3)$ m and the receiver at $(x_r,z_r)=(12,3)$ m.

In order to start the reference calculation, the pressure, velocity, density, and energy fields need to be specified. A Gaussian pulse is propagated using a one-dimensional version of the code presented in the Appendix. By adjusting the pulse amplitude and width one can obtain a one-dimensional signal at a given distance. In this example an amplitude and signal length of approximately 4 kPa and 1.5 m, respectively, were aimed for at a distance of 3 m from the source. Spatial steps are equal to 0.015 m in both directions, leading to a resolution of approximately 100 points per wavelength. This signal is then spherically extrapolated to obtain a 2D array. Figure 3 shows the one-dimensional signal and its 2D extension used to start both reference and NPE calculations.

A simulation on a perfectly rigid ground has been performed together with two calculations on different ground layers. Both have identical tortuosity ($\Phi=3$) and porosity ($\Omega_0=0.3$) but have different flow resistivity values ($\sigma_0=100$ kPa s m^{-2} and $\sigma_0=10$ kPa s m^{-2}). These flow resistivity values have been chosen to test the model limitations rather than to represent a real situation. Chosen flow resistivities would correspond to grass ($\sigma_0=100$ kPa s m^{-2}) and light, dry snow ($\sigma_0=10$ kPa s m^{-2}). The ground layer is 75 cm thick (50 points) and for NPE calculations the moving-window is 4.5 m wide and 6 m high (300×400 points).

3. Results

Figure 4 shows snapshots of the propagation for nonrigid ground layers at time $t=33$ ms for both models. Colormaps represent results from the NPE model while contour lines are results from Euler's equations. Time signals are recorded at the receivers; Fig. 5 shows these signals for NPE and reference calculations for the three ground layers considered. Although Euler's equations' model seems to smear out reflected waves more than the NPE model, the parabolic propagation model produces time waveforms comparable to the reference ones.

To evaluate the accuracy of the NPE model, some characteristics of the reflected wave are studied: the maximum positive and negative peak pressures and their arrival times (noted, respectively, p_+ and p_-, and t_{a+} and t_{a-}), and the positive phase duration (noted t_d). These characteristics are summarized in Table I. Since for the softest ground layer the negative peak on the reflected wave is very weak, values of p_- and t_{a-} for this layer are irrelevant.

As one can see, arrival times differ by at most 0.3 ms. The difference is larger for the softest layer; a possible reason is that the NPE model does not smear out pulses as the reference model does, leading to erroneous positive peak position. One can thus expect that as the flow resistivity decreases the error on arrival time increases. However, in outdoor sound propagation applications, the flow resistivity may seldom be lower than the one used here ($\sigma_0=10$ kPa s m^{-2}), so the error on arrival time will remain

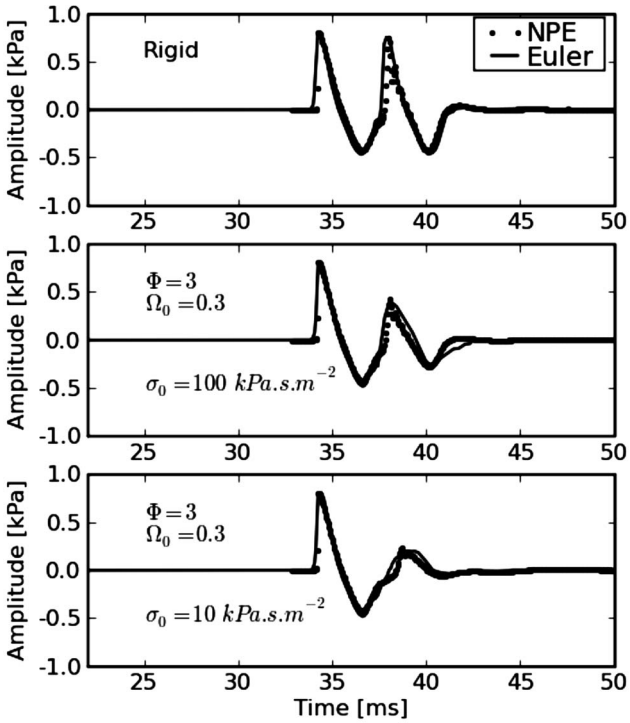


FIG. 5. Time signals at the receiver for NPE and reference calculations for the three ground layers considered (from top to bottom: perfectly rigid, $\sigma_0=100$ kPa s m^{-2} and $\sigma_0=10$ kPa s m^{-2}). Solid line: Euler; dotted line: NPE.

weak for most cases. These remarks are also applicable to the positive phase duration t_d. Positive peak amplitudes differ by 6.2% and 5.1% for layers with $\sigma_0=100$ kPa s m^{-2} and $\sigma_0=10$ kPa s m^{-2}, respectively. This difference does not seem to be dependent on flow resistivity and, as a comparison, the relative error for the perfectly rigid layer is 1%. Relative errors for negative peaks are comparable: 2.4% and 3.4% for the rigid layer and the layer with $\sigma_0=100$ kPa s m^{-2}, respectively.

As a mean of comparison, calculation times for Euler and NPE models were about 3.5 h and 4 min, respectively (calculations were done on a modern desktop computer). Although Euler's equations' implementation could use more advanced numerical techniques (adaptive mesh refinement methods[25] and moving-window principle[15]), the NPE model, thanks to the use of a single variable one-way wave equation and a fast solver (Thomas algorithm), is a very efficient tool for outdoor sound propagation simulations.

V. INCLUDING FORCHHEIMER'S NONLINEARITIES IN THE TWO-WAY COUPLING

While the flow resistivity dependence on particle velocity (Forchheimer's nonlinearities) is accounted for in the NPE model for porous ground layers [last term in Eq. (19)], the two-way coupling between both domains does not contain high-amplitude effects on ground properties. This would lead to erroneous solutions, since an additional attenuation would be introduced in the ground layer, but the increased rigidity of the interface would not be accounted for.

A solution is to artificially increase the flow resistivity in Eqs. (30a)–(30c) according to

TABLE I. Reflected wave characteristics for reference and NPE calculations.

Model	t_{a+} (ms)	p_+ (Pa)	t_d (ms)	t_{a-} (ms)	p_- (Pa)
		Rigid layer			
Euler	37.9	756	1.4	40.2	−449
NPE	38.0	749	1.3	40.0	−438
		σ_0=100 kPa s m^{-2}			
Euler	38.1	387	1.8	40.3	−265
NPE	38.2	411	1.6	40.2	−274
		σ_0=10 kPa s m^{-2}			
Euler	39.3	202	2.4	⋯	⋯
NPE	39	213	2.0	⋯	⋯

$$\sigma(x,t) = \sigma_0(1 + \xi w^i), \tag{39}$$

where w^i is the vertical particle velocity at the interface. Note that the flow resistivity is now dependent on (x,t); it is thus noted $\sigma(x,t)$. The authors then use Eq. (22) to obtain an approximation of w^i as follows:

$$w^i = (\rho_0 c_0)^{-1} \int \partial_z p_1^i dx + O(\epsilon^{5/2}), \tag{40}$$

where p_1^i is the first-order approximation of the pressure at the interface. The flow resistivity σ_0 in the coupling parameters (30a)–(30c) is thus replaced with

$$\sigma(x,t) = \sigma_0\left(1 + \frac{\xi}{\rho_0 c_0} \int \partial_z p_1^i dx\right). \tag{41}$$

At each time iteration the flow resistivity is updated with the help of pressure values at the interface at the previous time step. This method, although approximate, allows inclusion of Forchheimer's nonlinearities in the two-way coupling.

A. Numerical example

To illustrate the effects of Forchheimer's nonlinearities a simulation is performed with a nonlinearity parameter ξ =2.5 s m^{-1}. According to the conclusions of the previous section, low flow resistivities lead to a larger error on positive phase duration and time of arrival of positive peak pressure. A low flow resistivity has been chosen (σ_0 =10 kPa m s^{-2}, with ϕ=3 and Ω_0=0.3), so that the method to include Forchheimer's nonlinearities can be fully evaluated.

Simulation parameters and initialization array are identical to those used in Sec. IV B. The source is positioned at $(x_s,z_s)=(0,3)$ m and the receiver at $(x_r,z_r)=(12,3)$ m.

Figure 6 shows time signals at the receiver and Table II summarizes their characteristics for both reference and NPE calculations. One can see that compared to the same ground layer with no Forchheimer's nonlinearities (bottom plot in Fig. 5), the obtained reflected wave has a larger positive peak amplitude and a shorter time of arrival. The relative errors for positive and negative peak pressures are 4.38% and 4.34%, respectively, while the error on positive phase duration is 0.4 ms. These values, in agreement with the ones found in Sec. IV B, seem to indicate that the method used to include Forchheimer's nonlinearities in the two-way coupling does not introduce any additional source of error.

To confirm this statement, differences in signals characteristics for calculations with and without nonlinearities are studied. Table III presents these figures for both models. The positive peak amplitude is increased by 21.78% and 21.12%, and the time of arrival t_{a+} is reduced by 0.7 and 0.6 ms for Euler and NPE simulations, respectively, while the positive phase duration is reduced by 0.3 ms for both models. The signals modifications due to the addition of Forchheimer's nonlinearities are nearly identical for both models, confirming that the method presented to take into account the flow resistivity dependence on particle velocity is accurate.

VI. CONCLUSION AND PERSPECTIVES

A NPE model based on a nonlinear extension of the ZK model has been developed; it allows simulation of weakly nonlinear propagation within a porous ground layer. Next, two-way coupling equations have been derived from linearized Euler's equations. This interfacial boundary condition couples air and ground NPE models and enables the NPE to account for the effects of soft ground layers on sound propagation. For linear propagation, the results obtained with this method have shown very good agreement with analytical solutions for a wide range of ground properties. For high-

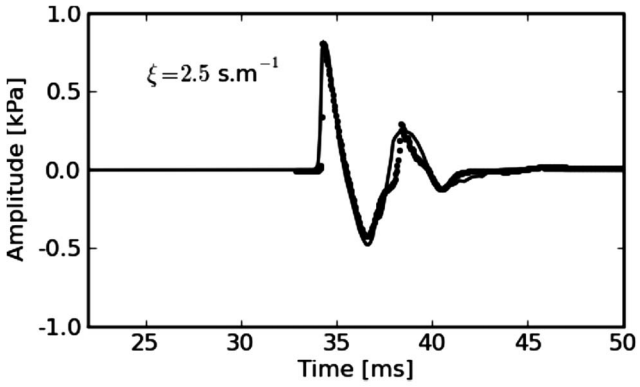


FIG. 6. Time signals at the receiver for Forchheimer's parameter ξ =2.5 s m^{-1}. Solid line: Euler; dotted line: NPE.

TABLE II. Reflected wave characteristics for reference and NPE calculations with Forchheimer's nonlinearities.

Model	t_{a+} (ms)	p_+ (Pa)	t_d (ms)	t_{a-} (ms)	p_- (Pa)
Euler	38.6	246	2.1	40.4	115
NPE	38.4	258	1.7	40.5	120

amplitude waves, the NPE model produces time signals comparable to those obtained by Euler's equations' model. Relative error on peak pressures has been shown to be independent of material properties while differences on positive phase duration and time of arrival increase with decreasing ground flow resistivity. However, the presented model still gives good agreement even for very low flow resistivities and provides a simple but efficient way of taking into account ground impedances. Finally, an approximate method to include Forchheimer's nonlinearities in the two-way coupling has been presented: It consists in artificially increasing the flow resistivity value in the coupling parameters. This method has been proven to give satisfactory results and does not introduce any additional source of error in the two-way coupling.

To construct the NPE model, the assumption that the ground layer is equivalent to a continuous fluid has been made. This simplified approach allows derivation of a ground model that is of the same form as the NPE model for atmospheric layer. Since the two-way coupling equations involve only spatial derivatives and integrals, the complete NPE model is able to perform simulations in a very short time (about 50 times faster than Euler's equations' implementation). This enables the NPE model to be used as a stochastic model solved by the Monte Carlo method. Wave field statistics in the air layer could be determined by performing a large number of simulations of sound propagation in an environment with varying parameters (e.g., propagation over a ground layer with random flow resistivity and propagation through turbulences). However, note that for realistic simulations, a spherical spreading term should be added to the NPE used in this work.

The relative simplicity of the NPE model and its coupling method makes it a good candidate for extensions and modifications. In a previous work,[35] the NPE models for porous ground layers and two-way coupling equations have been adapted to handle non-flat topographies, through the terrain-following coordinates method.[19] Two-way coupling equations could also be derived for multilayered ground surfaces without much additional work. With atmospheric refraction and dissipation included, it would provide a complete NPE model for weakly nonlinear wave propagation including most of the features of sound propagation outdoors (refraction, dissipation, topography, and ground impedance effects). This tool could be used, for instance, for propagating waves from explosions using a three stage procedure: first, a method based on Euler's equations could be used in the near field, where the propagation is highly nonlinear. Next, NPE models could propagate weakly nonlinear waves over moderate distances and, finally, when the wave amplitude is low enough, frequency-domain method such as the parabolic equation (PE) could be used. This hybrid method allows propagation of waves from explosions over distances up to several kilometers.[3]

TABLE III. Differences in reflected waves characteristics with and without Forchheimer's nonlinearities. Results are shown for both NPE and reference calculations.

Model	t_{a+} (ms)	p_+ (Pa)	t_d (ms)
Euler	−0.7	+21.78%	−0.3
NPE	−0.6	+21.12%	−0.3

ACKNOWLEDGMENTS

The authors would like to thank F. Van Der Eerden and E. Salomons for providing the initialization array and answering their questions about it.

APPENDIX: NUMERICAL SOLUTION OF EULER'S EQUATIONS

The main principle of the WENO scheme is the use of multiple stencils to evaluate the derivative at a given point. The algorithm first determines where there is a discontinuity and then weights stencils accordingly to avoid spurious numerical oscillations. These features make the WENO scheme accurate for propagating shock waves. For the sake of brevity computation details are omitted here, but the reader may refer to the work of Shu[33] or Wochner.[16,36]

The time discretization scheme is of the form

$$\mathbf{w}^{(1)} = \mathbf{w}^n + \Delta t \mathbf{K}^n, \tag{A1a}$$

$$\mathbf{w}^{(2)} = \tfrac{3}{4}\mathbf{w}^n + \tfrac{1}{4}\mathbf{w}^{(1)} + \tfrac{1}{4}\Delta t \mathbf{K}^{(1)}, \tag{A1b}$$

$$\mathbf{w}^{n+1} = \tfrac{1}{3}\mathbf{w}^n + \tfrac{2}{3}\mathbf{w}^{(2)} + \tfrac{2}{3}\Delta t \mathbf{K}^{(2)}, \tag{A1c}$$

where, for the air layer, $\mathbf{w}^n$ is the solution vector at time iteration n, i.e.,

$$\mathbf{w}^n = \begin{pmatrix} \rho_T \\ \rho_T u \\ \rho_T w \\ \rho_T e_0 \end{pmatrix}^n \tag{A2}$$

and $\mathbf{K}^{(i)}$ is the right hand side of the equation system, i.e.,

$$\mathbf{K}^{(i)} = -\partial_x \begin{pmatrix} \rho_T u \\ \rho_T u^2 \\ \rho_T uw \\ \rho_T e_0 u \end{pmatrix}^{(i)} - \partial_z \begin{pmatrix} \rho_T w \\ \rho_T uw \\ \rho_T w^2 \\ \rho_T e_0 w \end{pmatrix}^{(i)} - \begin{pmatrix} 0 \\ \partial_x p_T \\ \partial_z p_T \\ -\partial_x (p_T u) - \partial_z (p_T w) \end{pmatrix}^{(i)}. \quad \text{(A3)}$$

Note that for the ground layer, $\mathbf{w}^n$ and $\mathbf{K}^{(i)}$ have to be modified according to Eqs. (36a) and (36b).

Although the combination of WENO and Runge–Kutta schemes allow to stably propagate discontinuities, it is unable to propagate waves of infinite slope: A shock smearing will occur where the slope is too steep, resulting in small deviations from physical solutions for very high-amplitude waves.

[1] B. E. McDonald and W. A. Kuperman, "Time domain formulation for pulse propagation including nonlinear behaviour at a caustic," J. Acoust. Soc. Am. **81**, 1406–1417 (1987).
[2] K. Castor, P. Gerstoft, P. Roux, B. McDonald, and W. Kuperman, "Long-range propagation of finite-amplitude acoustic waves in an ocean waveguide," J. Acoust. Soc. Am. **116**, 2004–2010 (2004).
[3] F. van der Eerden and E. Védy, "Propagation of shock waves from source to receiver," Noise Control Eng. J. **53**, 87–93 (2005).
[4] K. Attenborough, P. Schomer, E. Védy, and F. van der Eerden, "Overview of the theoretical development and experimental validation of blast sound-absorbing surfaces," Noise Control Eng. J. **53**, 70–80 (2005).
[5] K. Attenborough, A. Cummings, P. Dutta, P. Schomer, E. Salomons, E. Standley, O. Umnova, F. van den Berg, F. van der Eerden, P. van der Weele, and E. Védy, "Blast sound absorbing surfaces. A joint project of the ERDC and the Netherlands ministry of defense," Technical Report, Engineer Research and Development Center/Cold Regions Research and Engineering Laboratory, Hanover, NH, 2004.
[6] T. Leissing, "Nonlinear outdoor sound propagation—A numerical implementation and study using the nonlinear progressive wave equation," MS thesis, Chalmers University of Technology, Göteborg, 2007.
[7] B. E. McDonald, P. Caine, and M. West, "A tutorial on the nonlinear progressive wave equation (NPE)—Part 1," Appl. Acoust. **43**, 159–167 (1994).
[8] B. E. McDonald, "Weak shock interaction with a free-slip interface at low grazing angles," J. Acoust. Soc. Am. **91**, 718–733 (1992).
[9] P. Caine and M. West, "A tutorial on the nonlinear progressive wave equation (NPE)—Part 2. Derivation of the three dimensional Cartesian version without use of perturbation expansions," Appl. Acoust. **45**, 155–165 (1995).
[10] G. P. Too and J. H. Ginsberg, "Cylindrical and spherical coordinate versions of the NPE for transient and steady-state sound beams," J. Vibr. Acoust. **114**, 420–424 (1992).
[11] B. E. McDonald, "High-angle formulation for the nonlinear progressive wave equation model," Wave Motion **31**, 165–171 (2000).
[12] G. P. J. Too and S. T. Lee, "Thermoviscous effects on transient and steady-state sound beams using nonlinear progressive wave equation models," J. Acoust. Soc. Am. **97**, 867–874 (1995).
[13] J. J. Ambrosiano, D. R. Plante, B. E. McDonald, and W. A. Kuperman, "Nonlinear propagation in an ocean acoustic waveguide," J. Acoust. Soc. Am. **87**, 1473–1481 (1990).
[14] A. A. Piacsek, "Atmospheric turbulence conditions leading to focused and folded sonic boom wave fronts," J. Acoust. Soc. Am. **111**, 520–529 (2002).
[15] V. W. Sparrow and R. Raspet, "A numerical method for general finite amplitude wave propagation in two dimensions and its application to spark pulses," J. Acoust. Soc. Am. **90**, 2683–2691 (1991).
[16] M. S. Wochner, A. A. Atchley, and V. W. Sparrow, "Numerical simulation of finite amplitude wave propagation in air using a realistic atmospheric absorption model," J. Acoust. Soc. Am. **118**, 2891–2898 (2005).
[17] D. Heimann and R. Blumrich, "A linearized Eulerian sound propagation model for studies of complex meteorological effects," J. Acoust. Soc. Am. **112**, 446–455 (2002).
[18] V. E. Ostashev, D. K. Wilson, L. Liu, D. F. Aldridge, N. P. Symons, and D. Marlin, "Equations for finite-difference, time-domain simulation of sound propagation in moving inhomogeneous media and numerical implementation," J. Acoust. Soc. Am. **117**, 503–517 (2005).
[19] R. Karle and D. Heimann, "A linearized Eulerian finite-difference time-domain sound propagation model with terrain-following coordinates," J. Acoust. Soc. Am. **119**, 3813–3821 (2006).
[20] E. M. Salomons, R. Blumrich, and D. Heimann, "Eulerian time-domain model for sound propagation over a finite impedance ground surface. Comparison with frequency-domain models," Acta. Acust. Acust. **88**, 483–492 (2002).
[21] D. K. Wilson, V. E. Ostashev, S. L. Collier, N. P. Symons, D. F. Aldridge, and D. H. Marlin, "Time-domain calculations of sound interactions with outdoor ground surfaces," Appl. Acoust. **68**, 173–200 (2007).
[22] D. K. Wilson, S. L. Collier, V. E. Ostashev, D. F. Aldridge, N. P. Symons, and D. H. Marlin, "Time-domain modeling of the acoustic impedance of porous surfaces," Acust. Acta Acust. **95**, 965–975 (2006).
[23] D. K. Wilson, V. E. Ostashev, and S. L. Collier, "Time-domain equations for sound propagation in rigid-frame porous media," J. Acoust. Soc. Am. **116**, 1889–1892 (2004).
[24] J. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves," J. Comput. Phys. **114**, 185–200 (1994).
[25] T. Plewa, *Adaptive Mesh Refinement—Theory and Applications: Proceedings of the Chicago Workshop on Adaptive Mesh Refinement Methods, Sept. 3–5, 2003* (Springer, Berlin, 2005).
[26] P. H. Forchheimer, ""Wasserbewegung durch boden" (Water movement through soil)," Z. Ver. Dtsch. Ing. **45**, 1782–1788 (1901).
[27] D. K. Wilson, J. D. McIntosh, and R. F. Lambert, "Forchheimer-type nonlinearities for high-intensity wave propagation of pure tones in air-saturated porous media," J. Acoust. Soc. Am. **84**, 350–359 (1988).
[28] C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, New York, 1949).
[29] A. Krylov, S. Sorek, A. Levy, and G. Ben-Dor, "Simple waves in saturated porous media (I. The isothermal case)," JSME Int. J., Ser. B **39**, 294–298 (1996).
[30] E. Védy, "Simulations of flows in porous media with a flux corrected transport algorithm," Noise Control Eng. J. **50**, 211–217 (2002).
[31] O. Umnova, K. Attenborough, and A. Cummings, "High amplitude pulse propagation and reflection from a rigid porous layer," Noise Control Eng. J. **50**, 204–210 (2002).
[32] X. Di and L. E. Gilbert, "An exact Laplace transform for a point source above a ground surface," J. Acoust. Soc. Am. **93**, 714–720 (1993).
[33] B. Cockburn, C. Johnson, C.-W. Shu, and E. Tadmor, "Essentially non-oscillatory schemes for hyperbolic conservation laws," *Advanced Numerical Approximation of Nonlinear Hyperbolic Equations*, Lecture Notes in Mathematics, edited by A. Quarteroni (Springer, Berlin, 1998), Vol. **1697**, pp. 325–432.
[34] *Numerical Recipes in Fortran*, edited by W. Press, B. Flannery, S. Teukolsky, and W. Vetterling (Cambridge University Press, Cambridge, 1996), Chap. 16, pp. 702–708.
[35] T. Leissing, P. Jean, J. Defrance, and C. Soize, "Nonlinear parabolic equation model for finite-amplitude sound propagation in an inhomogeneous medium over a nonflat, finite-impedance ground surface," in Proceedings of EuroNoise 08, Paris (2008), pp. 1889–1894.
[36] M. S. Wochner, "Numerical simulation of multi-dimensional acoustic propagation in air including the effects of molecular relaxation," Ph.D. thesis, The Pennsylvania State University, Pennsylvania (2006).

Identification of the aeroacoustic response of a low Mach number flow through a *T*-joint

P. Martínez-Lera and C. Schram
LMS International, Researchpark Z1, Interleuvenlaan 68, 3001 Leuven, Belgium

S. Föller, R. Kaess, and W. Polifke
Lehrstuhl für Thermodynamik, Technische Universität München, D-85747 Garching, Germany

(Received 30 January 2009; revised 5 June 2009; accepted 5 June 2009)

A methodology to study numerically the aeroacoustic response of low Mach number confined flows to acoustic excitations is presented. The approach combines incompressible flow computations, vortex sound theory, and system identification techniques, and is applied here to study the behavior of a two-dimensional laminar flow through a *T*-joint. Comparison with experimental results available in literature shows that the computed source models capture the main physical mechanisms of the sound production in the shear layer of the *T*-joint.
 [DOI: 10.1121/1.3159604]

PACS number(s): 43.28.Py, 43.28.Ra [AH] Pages: 582–586

I. INTRODUCTION

In a pipeline system, acoustic resonances can arise, typically for frequencies such that a characteristic length corresponds to an integer number of quarters of the acoustical wavelength. In this situation, the shear layer at *T*-joints can either absorb or generate sound, thus preventing a resonance from appearing or, on the contrary, contributing to its growth and persistence in time. The response to external excitations of the flow through a *T*-joint is investigated with numerical techniques in the present work.

A methodology is proposed combining incompressible flow simulations, vortex sound theory,[1] and system identification techniques that have been applied in the past for acoustic transfer matrix identification.[2,3] The proposed approach is suitable for low Mach number flows and acoustically compact configurations.

II. INCOMPRESSIBLE FLOW THROUGH A *T*-JOINT

Figure 1 illustrates the directions of the main flow through the *T*-joint, which enters the main pipe through inlet 1 and exits through outlet 2. There is no main flow entering or exiting through the side-branch inlet 3. The main pipe and the side-branch have the same section S and diameter D. The flow is characterized by the constant density ρ_0 and the mean flow velocity U. The Reynolds number, based on D and U, is Re=1000, although additional simulations are run with Re =2000 and Re=3000 to study the influence of the Reynolds number on the response of the shear layer.

Inlet 1 is placed upstream at a distance $3D$ from the *T*-joint, and characterized by a constant velocity profile of momentum thickness $\theta_{\text{inlet}}=0.0065D$. The momentum thickness is defined here as

$$\theta = \int_{y_w}^{y_c} \frac{u(y)}{U_c}\left(1-\frac{u(y)}{U_c}\right)dy, \tag{1}$$

where $u(y)$ is the x-velocity profile, U_c is the x-velocity at the centerline of the duct, and y_w and y_c are, respectively, the y-coordinates of the lower wall and of the centerline of the duct.

The side-branch length is $3D$, with inlet 3 described through a zero-velocity condition. Outlet 2 is placed $9D$ downstream from the *T*-joint, and the outlet condition consists in fixing the x-derivative of the velocities to zero. The remaining boundaries correspond to duct walls, and were characterized as no-slip conditions.

A finite volume commercial code was used. The chosen solver is time-dependent, implicit, and second-order in time and space. The flow model is incompressible and two-dimensional. No turbulence modeling was applied. The mesh contains around 200 000 quadrilateral cells. A study of grid independence showed that a mesh twice as coarse in each coordinate direction provides the same results. The flow was computed for a total time $t>50D/U$ to obtain an accurate unexcited initial flow condition. Although the solver was time-dependent, a steady solution was obtained when no external excitations were applied.

The momentum thickness θ at the upstream edge of the *T*-joint was calculated from the results of the unexcited simulations according to Eq. (1). It yielded $\theta=0.026D$ for Re =1000, $\theta=0.020D$ for Re=2000, and $\theta=0.017D$ for Re =3000. The influence of this parameter is discussed in Secs. IV and V.

A. Flow with small external perturbations

The flow is excited with incoming velocity perturbations. As the flow solver is incompressible, the perturbations propagate instantaneously throughout the whole domain, and behave therefore as an acoustic excitation propagating with an infinite speed of sound. This corresponds to the limit of zero Helmholtz number, and therefore implies acoustical compactness of the *T*-joint.

If the amplitude of the excitation signals is small enough, the response of the shear layer to perturbations is linear. Two plane uncorrelated broadband excitations are applied to the flow: u_1' is added at inlet 1 and u_3' at the side-

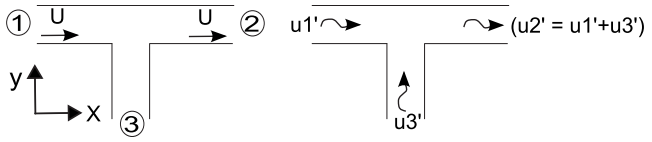


FIG. 1. *T*-joint: main flow and acoustic flow.

branch inlet 3. As the flow is incompressible, it is not possible to prescribe an additional perturbation at outlet 2, as the mass conservation in the joint defines it as a function of the other two: $u_2'=u_1'+u_3'$. The signals u_1' and u_3' are generated as discrete random binary signals, receiving at each timestep a value of either $-u_{bb}'$ or u_{bb}', and are then low-pass filtered to retain only a range of frequencies of interest. A typical value for u_{bb}' is $u_{bb}'=0.001U$. The total computation time of the excited flow is $t=80D/U$ with a timestep of $\Delta t=0.01D/U$.

B. Flow with moderate external perturbations

If the amplitude of the acoustic velocity perturbations in the *T*-joint becomes large enough compared to U, the response of the shear layer becomes nonlinear. In this situation, the flow must be separately analyzed for each frequency and direction of the acoustic flow, as done by Hofmans.[4] For each configuration shown in Fig. 2, simulations are carried out with single-frequency sinusoidal excitation signals of amplitudes $0.01U \leq u' < 0.5U$. The typical computation time is of around 12 periods of the excitation frequency, and the timestep is $\Delta t=0.01D/U$.

III. SOURCE PROCESSING

The area-averaged total pressure at several sections of the duct is stored at each timestep of the flow simulation. This allows computing the total pressure differences between the ducts separated by the *T*-joint: $\Delta P_{12}=P_1-P_2$ is the total pressure jump between positions x_1 in the inlet duct and x_2 in the outlet duct, while $\Delta P_{32}=P_3-P_2$ is the total pressure jump between certain positions y_3 in the side-branch and x_2 in the outlet duct. The sound production is obtained by using these flow data to compute equivalent sources of sound, which are defined according to vortex sound theory.[1]

For an incompressible fluid with sufficiently high Reynolds number, the momentum equation yields

$$\nabla P = -\rho_0 \frac{\partial \mathbf{u}}{\partial t} - \rho_0(\boldsymbol{\omega} \times \mathbf{u}), \tag{2}$$

where P is the total pressure $P=p+\rho_0 u^2/2$ and $\omega=\nabla\times\mathbf{u}$ represents vorticity. The second term on the right-hand side of Eq. (2) is related to the acoustic power W according to Howe's energy corollary,[1] which expresses the acoustic energy radiated in a control volume V as

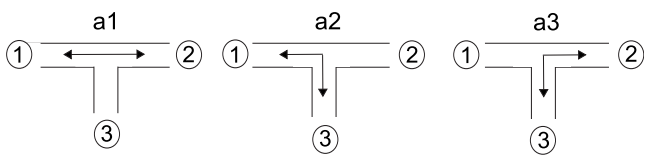


FIG. 2. Acoustic flow in each configuration.

$$W = -\rho_0 \int_V (\boldsymbol{\omega} \times \mathbf{u}) \cdot \mathbf{u}' dV. \tag{3}$$

The flow in the ducts is considered one-dimensional. In view of Eqs. (2) and (3), it can be deduced that the total pressure difference ΔP between two duct sections is the sum $\Delta P = \Delta p_{\text{pot}} + \Delta p_s$ of two components:[4] a pressure difference due to the sources of sound Δp_s, which is related to the vortical disturbances in the shear layer, and a pressure difference Δp_{pot} related to a potential flow solution, which can be expressed as

$$\Delta p_{\text{pot}} = -\rho_0 \int_L \frac{\partial u_l}{\partial t} dl, \tag{4}$$

where L is the total length between the duct sections, l is the coordinate along the duct, and u_l is the velocity component in its direction. Δp_{pot} may be interpreted as a result of the propagation along the duct of the fluctuations in the velocity u_l.

In order to quantify the sound production, the source pressure difference Δp_s must be calculated from the total pressure difference ΔP measured in the flow. If the flow in a duct is incompressible and the duct has a constant section, the velocity does not vary with the coordinates along the duct, and its time derivative can be placed outside the integral in Eq. (4). Therefore, for two duct sections a and b with a joint between them, the total pressure jump can be expressed as

$$\Delta P_{ab}(x_a,x_b,t) = c_a(t)L_a(x_a) + c_b(t)L_b(x_b) + (\Delta p_s)_{ab}(t), \tag{5}$$

where $c_a(t)$ and $c_b(t)$ are time-dependent coefficients that are proportional to the time derivatives of the velocities along the ducts, and L_a (L_b) is the total length of propagation along duct a (b)

$$L_a(x_a) = x_a - x_a^o (L_b(x_b) = x_b - x_b^o) \tag{6}$$

between the measuring position x_a (x_b) in the duct and a certain origin x_a^o (x_b^o), which is located somewhere inside the joint in order to include the effect of propagation inside it.

As a first approximation, it can be assumed that $\Delta p_{\text{pot}} \approx \Delta P$, as the source pressure jump Δp_s is small in comparison (typically, of the order of 1%). The total pressure in the duct is obtained by calculating the area-averaged pressure over the duct section, as the flow is considered one-dimensional. The measuring sections must be far enough from the joint, so that the flow is closer to a one-dimensional flow. In the present computations, the pressure was measured always at least a diameter D far from the joint. The local time-averaged value is subtracted from each pressure measurement. The total pressure jumps ΔP_{12} and ΔP_{32} are obtained for different measuring positions x_1 in the inlet duct, x_2 in the outlet duct, and y_3 in the side-branch. As a result, for each timestep t_k (with $k \in [1,N]$, N being the number of timesteps), linear least-squares fits of the total pressure jumps with respect to the measuring positions x_1, x_2, and y_3 can be carried out as

$$\Delta P_{12}(x_1,x_2,t_k) \approx c_1^k x_1 + c_2^k x_2 + c_0^k, \tag{7}$$

$$\Delta P_{32}(y_3,x_2,t_k) \approx \hat{c}_3^k y_3 + \hat{c}_2^k x_2 + \hat{c}_0^k. \tag{8}$$

The outputs of these linear fits are c_1^k, $\hat{c}_3^k$, c_2^k, $\hat{c}_2^k$, c_0^k, and $\hat{c}_0^k$, which are the coefficients that best approximate Eqs. (7) and (8) for the measured data. The next step is to compute the origins x_1^{o1}, x_2^{o1}, x_2^{o2}, and y_3^{o2} such that equations

$$\mathbf{c}_0 = -\mathbf{c}_1 x_1^{o1} - \mathbf{c}_2 x_2^{o1} \quad (\text{with } \mathbf{c}_i = \{c_i^k\}_{k=1}^N) \tag{9}$$

and

$$\hat{\mathbf{c}}_0 = -\hat{\mathbf{c}}_3 y_3^{o2} - \hat{\mathbf{c}}_2 x_2^{o2} \quad (\text{with } \hat{\mathbf{c}}_i = \{\hat{c}_i^k\}_{k=1}^N) \tag{10}$$

are satisfied in the sense of a least-squares approximation. The source pressure jumps can then be computed for each timestep from the total pressure measured at certain positions $x_1=\bar{x}_1$, $x_2=\bar{x}_2$, and $y_3=\bar{y}_3$ as

$$(\Delta p_s)_{12}(t_k) = \Delta P_{12}(\bar{x}_1,\bar{x}_2,t_k) - c_1^k(\bar{x}_1 - x_1^{o1}) - c_2^k(\bar{x}_2 - x_2^{o1}) \tag{11}$$

and

$$(\Delta p_s)_{32}(t_k) = \Delta P_{32}(\bar{y}_3,\bar{x}_2,t_k) - \hat{c}_3^k(\bar{y}_3 - y_3^{o2}) - \hat{c}_2^k(\bar{x}_2 - x_2^{o2}). \tag{12}$$

The source processing must be numerically robust, as the sources correspond only to a small fraction of the total pressure jump. This method based on linear fits with respect to the measuring coordinates proved more numerically robust than others based on the numerical time derivatives of the velocities in the ducts.

IV. ACOUSTIC RESPONSE IN THE LINEAR REGIME

The objective of the transfer function identification is to obtain a frequency-dependent matrix **A** that relates the output source pressure differences $(\Delta p_s)_{12}$ and $(\Delta p_s)_{32}$ to the input excitation velocities u_1' and u_3'.

$$\begin{pmatrix} (\Delta p_s)_{12} \\ (\Delta p_s)_{32} \end{pmatrix} = \rho_0 U \begin{pmatrix} A_{11}(\mathrm{Sr}) & A_{12}(\mathrm{Sr}) \\ A_{21}(\mathrm{Sr}) & A_{22}(\mathrm{Sr}) \end{pmatrix} \begin{pmatrix} u_1' \\ u_3' \end{pmatrix}, \tag{13}$$

where $\mathrm{Sr}=fD/U$ is the Strouhal number related to the excitation frequency f. The method applied to obtain the matrix is based on the Wiener–Hopf equation with multiple inputs.[2] It may be noted that, in order to complete the characterization of the relationship between acoustic inputs and outputs in the T-joint, for which three equations are necessary, the additional equation given by the mass conservation in the joint is needed: $u_2'=u_1'+u_3'$.

For each of the configurations described in Fig. 2, an impedance Z can be defined as follows.

- Configuration a1 $(u_1'=u_2';u_3'=0)$:

$$Z_{a1} = (\Delta p_s)_{12}/(\rho_0 U u_1') = A_{11}.$$

- Configuration a2 $(u_1'=-u_3';u_2'=0)$:

$$Z_{a2} = (\Delta p_s)_{13}/(\rho_0 U u_1') = ((\Delta p_s)_{12} - (\Delta p_s)_{32})/(\rho_0 U u_1') = A_{11} - A_{12} - A_{21} + A_{22}.$$

- Configuration a3 $(u_3'=u_2';u_1'=0)$:

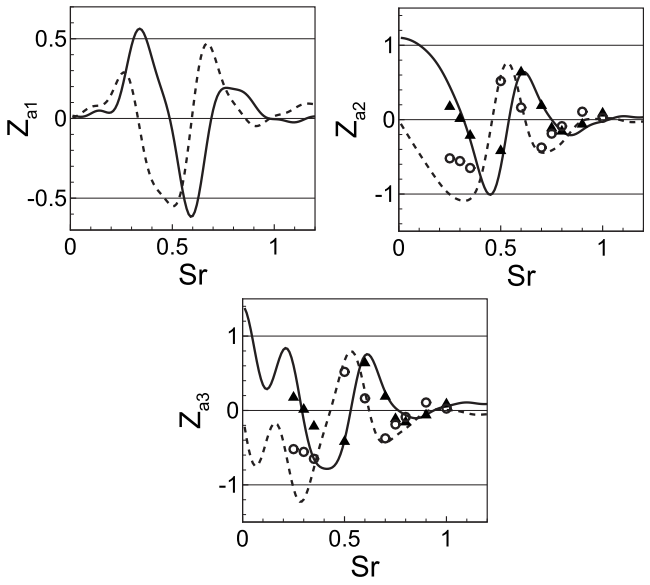


FIG. 3. Impedance for configurations a1, a2, and a3 vs Strouhal number: real (solid) and imaginary (dashed). The symbols correspond to experimental results by Graf and Ziada (Ref. 5)with $u'=0.014U$: real (▲) and imaginary (○).

$$Z_{a3} = (\Delta p_s)_{32}/(\rho_0 U u_3') = A_{22}.$$

In Fig. 3, the equivalent impedance obtained for the three configurations is shown. The impedance for configurations a2 and a3 is compared to experimental results obtained by Graf and Ziada,[5] who carried out measurements in a setup with two opposite coaxial side-branches. It must be stressed that the conditions of the experiments do not match exactly those of the numerical model. In the experiments, the acoustic flow is directed from one side-branch into another. This does not correspond to any of the configurations studied here, although it may be expected to behave similarly to a combination of configurations a2 and a3. Moreover, the amplitude of the experimental data shown in Fig. 3, the lowest for which experimental results are provided, corresponds to $u'=0.014U$, for which the shear layer may present already some nonlinearity. The diameter of the main duct in the experimental setup is 1.75 times the side-branch diameter. No information is available related to the experimental boundary layer profile at the upstream edge of the T-joint. Finally, some impact may be expected from the fact that the numerical simulations are two-dimensional and have a much lower Reynolds number.

All these differences make a quantitative comparison difficult. Nevertheless, from the comparison in Fig. 3, it can be concluded at least that the numerical approach provides a reasonable model of the sound production. The dependence with respect to the Strouhal number is correctly captured, as well as the frequency shift between real and imaginary parts, which also appears in experimental measurements of the acoustic impedance of an orifice with grazing flow.[6] Finally, the order of magnitude of the sources seems also realistic.

A direct relationship can be found between the impedances Z_{ai} and the average acoustic power $\langle W_{ai} \rangle$ as follows:

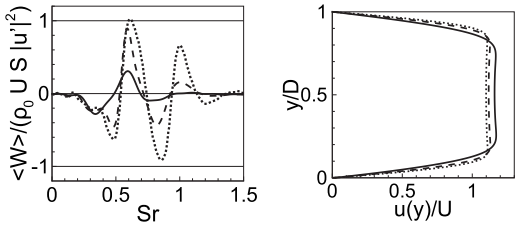


FIG. 4. Normalized average acoustic power for configuration a1 vs Strouhal number (left). On the right, the x-velocity profiles at the upstream edge of the T-joint are shown. Re=1000 (solid), Re=2000 (dashed), and Re=3000 (dotted).

$$\langle W_{ai}\rangle = \langle \mathrm{Re}(-\Delta p_s u' S)\rangle = -\tfrac{1}{2}\mathrm{Re}(Z_{ai})\rho_0 U S|u'|^2,$$

where $\langle\cdot\rangle$ denotes time average and ai indicates the configuration (a1, a2, or a3). The average acoustic power indicates whether sound will be on average generated ($\langle W_{ai}\rangle > 0$) or absorbed ($\langle W_{ai}\rangle < 0$), or, in other words, whether the shear layer will tend to amplify and sustain the perturbations, or rather to attenuate the acoustic field.

Figure 4 shows the average acoustic power for configuration a1. The results of additional computations with higher Reynolds numbers are shown too in order to illustrate the influence of different velocity profiles at the T-joint upstream edge, which are also shown in the figure. Kooijman *et al.*[6] reported that, for similar boundary layer profiles at the upstream edge of a rectangular orifice with grazing flow, the amplitude of the response increases with decreasing boundary layer thickness. The same trend is observed here. Furthermore, they also observed that above a Strouhal number based on the momentum thickness $\mathrm{Sr}_\theta = \mathrm{Sr}\theta/D$ the sources vanish, and that, therefore, the ratio θ/D determines the number of peaks observed in the acoustic impedance as a function of the Strouhal number Sr. The results shown in Fig. 4 are also consistent with this. It can be observed that, as the Reynolds number increases, the peak of sound production at Sr=0.6 becomes stronger, and a peak at Sr=1 appears. Whistling of modes corresponding to configuration a1 at around Sr=0.6 has been measured experimentally in a multiple side-branch system.[7]

Two tests were carried out in order to confirm that the system behaves linearly for small enough amplitudes: a test of the proportionality of the response with respect to the amplitude of the excitation and a test of the validity of the superposition principle with respect to excitations at the main inlet 1 and at the side-branch inlet 3.

V. ACOUSTIC RESPONSE IN THE NONLINEAR REGIME

The nonlinear regime is investigated by processing the sources for each flow simulation as explained in Sec. III. As the excitations for the nonlinear regime study contain a single frequency, there is no need to apply specific system identification techniques. Instead, the impedance for each simulation is obtained by dividing the Fourier coefficients of source pressure jump and acoustic velocity at the excitation frequency.

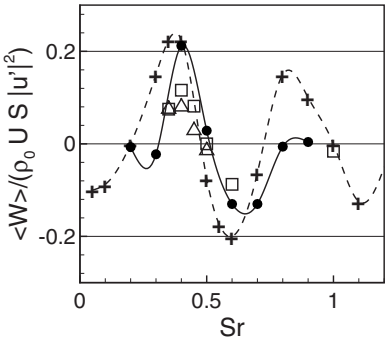


FIG. 5. Normalized average acoustic power vs Strouhal number for configuration a3 and $u'=0.2U$: present results (●) and Hofmans' results (+); experimental results by Graf and Ziada (Ref. 5): $u'=0.14U$ (□) and $u'=0.25U$ (△).

Figure 5 shows results for an acoustic excitation of amplitude $u'=0.2U$ and corresponding to configuration a3, together with those obtained by Hofmans[4] with an inviscid two-dimensional vortex blob method and the same amplitude $u'=0.2U$. The ratio $\langle W\rangle/|u'|^2$ tends to be smaller in absolute value for the nonlinear regime than for the linear regime. This is consistent with the idea that for larger amplitudes the shear layer tends to saturate. Experimental results obtained by Graf and Ziada[5] with $u'=0.14U$ and $u'=0.25U$ are also shown. All numerical and experimental curves predict the peak of maximum acoustic production around Sr=0.4. However, the experimental results show a peak that is lower than the one predicted by Hofmans[4] and by the present results. In any case, as explained in Sec. IV, the differences between the numerical and experimental setups can account for mismatches in the results, so it is difficult to extract conclusions from this.

Figure 6 illustrates the saturation of the response as the acoustic amplitude grows. It shows the ratio $\langle W\rangle/|u'|^2$, scaled so that it yields 1 for the smallest available acoustic amplitude, which corresponds to the linear regime for the numerical results and to $u'=0.014U$ for the experimental results by Graf and Ziada.[5] The given average acoustic power corresponds to the Strouhal number for which it is maximum (around Sr≈0.6 for configuration a1 and around Sr≈0.4 for

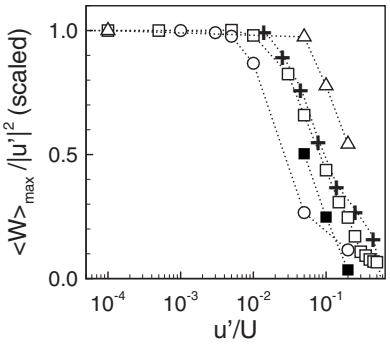


FIG. 6. Maximum average acoustic power vs acoustic amplitude: configuration a1 with Re=1000 (□), configuration a1 with Re=2000 (○), configuration a2 with Re=1000 (■), configuration a3 with Re=1000 (△), and experimental results by Graf and Ziada (Ref. 5) (+). The curves are scaled to obtain a value of 1 for the smallest available amplitude u'.

configurations a2 and a3, although in all cases it varies somewhat with the acoustic amplitude and the flow parameters).

For small amplitudes of the velocity, the shear layer behaves linearly, and, therefore, the power grows quadratically with u'/U, making the ratio $\langle W\rangle/|u'|^2$ constant. However, above a certain amplitude, the ratio $\langle W\rangle/|u'|^2$ starts to decay. For a range of amplitudes just above the linear range, it is approximately linear with $\log(u'/U)$. All presented curves, including numerical results for different configurations and Reynolds numbers and the experimental results, present similar decay rates, the main difference between them being the amplitude for which the decay starts. For different acoustic configurations, the onset of nonlinearities takes place at different values of the acoustic amplitude.

Kooijman *et al.*[6] reported that the onset of nonlinearities in an orifice with grazing flow takes place at lower acoustic amplitudes for decreasing boundary layer thickness. The same result is observed here, where the decay of $\langle W\rangle/|u'|^2$ for configuration a1 starts at an amplitude u' that is lower for Re=2000 than for Re=1000.

VI. FINAL CONSIDERATIONS

A method to characterize numerically the aeroacoustic behavior of low Mach number flows inside ducts has been established, and applied to investigate the sound production in an acoustically compact *T*-joint. The obtained results demonstrate the suitability of the proposed approach to study the response of low Mach number flows to acoustic excitations. The method differs from the usual applications of transfer matrix identification techniques in the fact that the flow simulation is incompressible, and the equivalent sources are defined through pressure differences (dipoles) using an acoustic analogy based on vortex sound theory.

The extent to which the obtained results are quantitatively representative of three-dimensional flows in real configurations still remains an open question. Two-dimensional disturbances tend to be in general more unstable than three-dimensional ones.[8] Further investigation is needed to clarify this.

ACKNOWLEDGMENTS

The authors are grateful to the EC for the support for the present work, performed within the Marie Curie RTN project AETHER (Contract No. MRTN-CT-2006-035713). Financial support by the Deutsche Forschungsgemeinschaft for R.K. (Grant No. Po 710/3) and S.F. (Grant No. Po 710/5) is also gratefully acknowledged.

[1] M. S. Howe, *Theory of Vortex Sound*, Cambridge Texts in Applied Mathematics (Cambridge University Press, Cambridge, 2003).
[2] W. Polifke, A. Poncet, C. O. Paschereit, and K. Döbbeling, "Reconstruction of acoustic transfer matrices by instationary computational fluid dynamics," J. Sound Vib. **245**, 483 (2001).
[3] S. Föller, R. Kaess, and W. Polifke, "Reconstruction of acoustic transfer matrices from large-eddy-simulations of complex turbulent flows," AIAA Paper No. 2008-3046 (2008).
[4] G. C. J. Hofmans, "Vortex sound in confined flows," Ph.D. thesis, Eindhoven University of Technology, Eindhoven, The Netherlands (1998).
[5] H. R. Graf and S. Ziada, "Flow induced acoustic resonance in closed side branches: an experimental determination of the excitation source," ASME Symposium on Flow-Induced Vibration and Noise (1992), AMD-Vol. 151/PVP-Vol. 247, Vol. **7**.
[6] G. Kooijman, A. Hirschberg, and J. Golliard, "Acoustical response of orifices under grazing flow: Effect of boundary layer profile and edge geometry," J. Sound Vib. **315**, 849 (2008).
[7] G. Nakiboglu, S. P. Belfroid, D. Tonon, J. F. H. Willems, and A. Hirschberg, "A parametric study on the whistling of multiple side branch system as a model for corrugated pipes," ASME Paper No. PVP2009-77754 (2009).
[8] A. V. Dovgal, V. V. Kozlov, and A. Michalke, "Laminar boundary layer separation: Instability and associated phenomena," Prog. Aerosp. Sci. **30**, 61 (1994).

Coherent and incoherent scattering by a plume of particles advected by turbulent velocity flow

David R. Palmer
Atlantic Oceanographic and Meteorological Laboratory, National Oceanic and Atmospheric Administration, 4301 Rickenbacker Causeway, Miami, Florida 33149

(Received 11 December 2008; revised 28 May 2009; accepted 30 May 2009)

Studies of acoustic remote sensing of the plumes that result from the injection of particulate matter in the ocean, either naturally or by dumping or dredging activities, have assumed the scattering is incoherent. These plumes are always turbulent, however. The particle density is a passive scalar that is advected by the turbulent velocity flow. The possibility exists, therefore, that the scattered waves from a significant number of particles add coherently as a result of Bragg scattering. In this paper, we investigate this possibility. We derive an expression for the ratio of the coherent intensity to the incoherent one in terms of the turbulent spectrum and the properties of the particles that make up the plume. The sonar is modeled as a high-Q, monostatic, pulsed sonar with arbitrary pulse envelope and arbitrary, but narrow, beam pattern. We apply the formalism to acoustic remote sensing of black smoker hydrothermal plumes. We find that, at most, the coherent intensity is less than 1% of the incoherent one. The implications are that Bragg scattering does not lead to a significant coherent component and in analyses of scattering from this type of plume, one can ignore the complications of turbulence altogether. [DOI: 10.1121/1.3158933]

PACS number(s): 43.30.Gv, 43.20.Fn, 43.30.Vh [WLS] Pages: 587–598

I. INTRODUCTION

The relative importance of coherent and incoherent acoustical scattering from a plume of inhomogeneities is a topic that dates to the time of Rayleigh.[1] The next notable advance after Rayleigh was probably in 1945 when Foldy[2] published an influential paper in which he considered scattering of a scalar wave from a collection of scatterers and derived equations satisfied by the coherent field and the intensity of the incoherent field. These equations are similar to free-field equations except the wave number is replaced by a spatially-dependent effective wave number depending on the cross section for a single scatterer and the geometry of the scattering situation. Foldy's work was extended by a number of authors including Lax,[3,4] Waterman and Truell,[5] and Ishimaru.[6] It is the basis of the analysis of coherent and incoherent plane wave scattering from a cloud of scatterers in the long wavelength or Rayleigh scattering approximation presented by Morse and Ingard.[7]

From the mid-1950s to the mid-1960s a number of papers were published[8–12] investigating the significance of coherent versus incoherent scattering in an effort to understand the oceanic deep scattering layer and the nature of acoustic signals scattered from it.[13] For the most part, these papers considered plane or spherical waves scattered from well-defined geometric regions (e.g., horizontal layers) in which the distribution of monopole scatterers is random but statistically uniform.

In recent years, interest in the relative importance of coherent and incoherent scattering stems primarily from the realization that oceanic bubble clouds can significantly scatter acoustic signals. Most of the work on scattering from a cloud of bubbles assumes a random but statistically uniform distribution of scatterers and focuses on the complications that result from the resonant structure of the single-scatterer cross section and from multiple scattering. Multiple scattering is usually taken into account using some variation of the effective medium approach pioneered by Foldy. The literature is too extensive to cite. However, the book by Leighton[14] contains citations too much of the literature.

From all this work, one reaches the general conclusion that there are two mechanisms that can lead to a significant coherent contribution to the intensity backscattered from a collection of scatterers in the ocean. First, the density of scatterers can be non-uniform. In the extreme the scatterers could be uniformly arranged on a grid. Bragg scatterering, well-known from x-ray diffraction by a crystalline structure, would then lead to a coherent component,[15] even in the single scattering approximation. Second, the density of scatterers is so large that the mean distance between them is of the same order as or smaller than the carrier wavelength of the sonar. In this case multiple scattering is important and the plume begins to act as a unified whole. The phase cancellations that characterize scattering from a collection of randomly distributed scatterers no longer take place and one can have, again, a significant coherent contribution. The need to account for multiple scattering and the presence of a coherent component go hand-in-hand.

We are interested in acoustic remote sensing, using high-frequency pulsed sonars, of the plumes of scatterers that result from injection into the ocean of particulate matter. The injection can be a natural phenomenon[16] or a man-made one.[17] The material can be solid particles, bubbles, or oil droplets in the case of a naturally occurring plume and dredge material, sewage sludge, or other potential marine pollutants in the case of a dumping or discharge activity. In all cases we assume scattering is in the long-wavelength or Rayleigh scattering region. The Rayleigh scattering ampli-

tude contains both monopole and dipole terms and the dipole term cannot be ignored. This is a significant complication for multiple scattering but not for single scattering.

A priori, there is no reason to believe that both mechanisms, a spatially non-uniform density and a large density, do not lead to a significant coherent component. The underwater plumes of interest to us are always turbulent. The particle density is a passive scalar advected by the turbulent velocity flow. Simple arguments, to be discussed, indicate the Bragg wavenumber is within the inertial subrange. Further, the estimated density of particles in the core of a black smoker hydrothermal plume is 30 particles/cm^3, giving a mean separation between particles of about 3 mm to be compared with a wavelength of about 1 cm for a typical plume imaging sonar. Plumes resulting from the dumping of material into the ocean are typically even more dense. There is reason to believe, therefore, that multiple scattering is also important and that the backscattered intensity has a significant coherent component.

The significance of a coherent component relates to the value of acoustic remote sensing. In the analysis[19] that was done to support the experimental work[19,20] imaging black smoker hydrothermal plumes, multiple scattering was not taken into account even though we recognized that it would likely be important. As a result, the validity of the analysis was stated to be limited to estimating the backscattered intensity from the boundary area of the plume rather than from the interior of the plume where densities are greater. Knowing the boundary of a plume is important for guiding chemical and physical sampling and determining the relationship between the plume and its ocean environment. At the time, these were considered the most important reasons for acoustically imaging hydrothermal plumes. Subsequently, an effort was undertaken[21,22] to use the backscattered intensity data to construct a three-dimensional image of the plume, including its interior, based on the phenomenological description of the mean density of particles within a turbulent, buoyant plume.[23] This effort assumes a number of additional assumptions most important of which is that the backscattered intensity is proportional to the particle density multiplied by the volume being ensonified by the sonar. If there is a significant coherent component, this assumption is not valid because then the backscattered intensity would be equal to the sum of two terms; one proportional to the particle density and another proportional to the particle density squared.

One might ask if it is possible to experimentally determine if a coherent component is present. In principle, it is possible. However, from the work done imaging hydrothermal plumes, it is highly unlikely this determination can be made. One reason is that we did not initially appreciate the importance of averaging over a large number of pings. As mentioned, the initial emphasis was on defining the boundary of a plume and plumes are so particle rich that it does not take much ping averaging to do this. However, estimating the average particle density requires high-quality estimates of the average intensity. For incoherent scattering, to obtain a 10% precision in the average intensity requires averaging 100 pings.[24] This level of precision was never obtained or even strived for. In addition, in the experimental work done on the Juan de Fuca Ridge,[20] the sonar transmitter and receiver were co-located but had different beam patterns—one consisted of a fan in the vertical and the other a fan in the horizontal. While this provided a sharp pencil beam, it also introduced noise from side lobes that would have been supressed with a true monostatic sonar where the intensity is proportional to the square of a single beam pattern.

Because of these considerations, it seems worthwhile to undertake an analysis to determine the significance of coherent scattering for plume imaging resulting from both turbulence and from multiple scattering. In this paper, we consider the turbulence problem. It is the obvious first step since it involves only single scattering and is preliminary to any analysis of multiple scattering.

We have tried to be as realistic as possible in modeling the sonar system so that our analysis would more likely be relevant to experimental situations than if it were based on a simple incident plane wave or spherical wave. The sonar emits a pulse of sound at a carrier frequency f_0 of duration τ, where $Q \equiv f_0 \tau \gg 1$. The envelope of the pulse is arbitrary. The pulse is transmitted and received in a narrow beam described by a beam pattern D. We assume the received wave is demodulated to base and its quadrature components obtained. From these components, the received complex demodulated pressure wave is obtained. In an actual processor, demodulating and constructing the quadrature components would require processing of a time interval of the received pressure. Here we simulate this processing by averaging the complex demodulated received wave over the duration of a pulse.

The analysis in this paper is characterized by two concepts that we find useful. First, in Appendix A, the sonar transmitter is included in the analysis not as a vibrating portion of the boundary of the ocean but rather as a source term in the wave equation. This source term, or volume density, is then restricted to a compact, planar surface—the active element of the transmitter—by using a delta function. We find this approach to be quite simple. Also, in the event one wants to consider a source consisting of a collection of point sources it is particularly convenient. Of course, both approaches in the end give equivalent results.

Second, and more importantly, we make extensive use of the concept of an *ensonified volume*, $\mathbb{V}$. The pressure recorded at the receiver at a particular time after the transmission of a pulse consists of the sum of the scattered waves from all the particles in the ensonified volume as it existed at an earlier time (allowing for the times it takes for the scattered waves to travel from the scatterers to the receiver). The ensonified volume depends on the time of recording, the beam patterns of the transmitter and receiver, the pulse characteristics of the transmitter as well as the orientation of the transmitter, i.e., the direction in which it points. In this paper, we use a precise mathematical definition of the ensonified volume that takes into account the fact its boundaries are not sharp and that the beam pattern has side lobes that result in the ensonification of ocean regions out of the main beam. Quantities are expressed as integrals over the ensonified volume and we give an expression for the size (in m^3) of this volume. For high-Q systems, the dimensions of the ensoni-

fied volume are large compared to the wavelength of the carrier frequency. If one integrates the phase term $\exp(ik_0 r)$, where k_0 is the carrier wavenumber, over the ensonified volume

$$\int_{\mathbb{V}} d\boldsymbol{r} \exp(ik_0 r),$$

one obtains essentially zero since one is integrating a phase term over many cycles of its variation. This characteristic is central to the discussion of incoherent versus coherent scattering and is used over and over in the paper.

As mentioned, we assume the dimensions of the particles that comprise the plume are so small in comparison to the wavelength of the carrier frequency that scattering is in the Rayleigh scattering regime. This is certainly true for hydrothermal plumes where the metallic sulfides that comprise the plume have linear dimensions less than about 35μ.[25,18] It is also true for many dredged material plumes.[17] The particles in the plume are assumed to be spherical, to support only compressional waves, and to have identical mechanical properties. For hydrothermal plumes, these assumptions have all been discussed in previous publications.[19,26,27] We assume the motion of the particles is such that they can be considered stationary during the time of passage of the sonar pulse.

In the analysis there are two types of averaging that are done. In Sec. II, a configuration averaging is done and then in Sec. IV an average over the turbulent fluctuations in the ocean media is carried out. (There is actually a third type of averaging—the running average of the recorded intensity over a period of time equal to the pulse length [Eq. (A21)]. This averaging, however, is simply technical and not of particular interest.) The configuration averaging is an average over all the possible positions in space of the N particles that make up the plume, as well as their possible volumes. This type of average is well-known from the work by Foldy.[2] After the configuration averaging, the scattering is described in terms of a particle density—a continous function of position. So the configuration averaging takes one from a description in terms of discrete scatterers to one in terms of a continous function of position. To describe the effects of turbulence, one then assumes that the particle density is a scalar field advected by the turbulent flow. This second type of averaging is then an average over the turbulent fluctuations. This analysis in terms of two types of averaging is somewhat artificial. Regardless of how well a description in terms of a density function may describe an experimental situation, it remains an approximation since scattering is always from decrete scatterers. It is the individual scatterers that are advected by the turbulent flow and not the density. In addition, the purpose of the analysis is to model an experimental situation and in an actual experiment the only averaging that is done is the averaging over a large number of pings. We have chosen to use a description in terms of two types of averaging because it allows us to better make connection with previous work; particularly the work by Foldy[2] and the large body of research that followed from his work.

This paper is organized as follows. In Sec. II, we obtain expressions for the coherent and incoherent intensities as integrals, over a continuous particle density by using configuration averaging as was done by Foldy.[2] In Sec. III, we define the ensonified volume and discuss some general characeristics of the coherent and incoherent intensities. In Sec. IV, we assume the scatterers are advected by turbulent fluctuations in the flow and develop an expression for the ratio of the coherent intensity to the incoherent one. This expression is further developed by assuming the Bragg wavelength is within the inertial subrange and then using a model for the turbulent spectrum in this subrange. Section V is devoted to an application of the analysis to black smoker hydrothermal plumes. In Sec. VI, we have a brief discussion of results. In Appendix A, starting from first principles, we derive an expression for the complex demodulated pressure recorded by a monostatic sonar. We have included this preliminary material not only to establish notation and to make the paper self-contained, but also because we did not find it available elsewhere in a form that fits our needs.[28] Appendix B is devoted to a discussion of the structure of an integral that appears in the analysis.

II. CONFIGURATION AVERAGING

In Appendix A we derive the following expression for the complex demodulated pressure $P(t)$, recorded at time t by a monostatic sonar and averaged over a time equal to the pulse length due to a pulse transmitted at time t_0:

$$P(t) = \mathbb{C}(t) \sum_i v_i D^2(\hat{r}_i) W(r_i, t) e^{2ik_0 r_i}, \tag{1}$$

where v_i is the volume of the ith particle in the plume, $\boldsymbol{r}_i\,(=r_i\hat{r}_i)$ is the position vector from the center of the transmitter to the location of the ith particle, $D(\hat{r}_i)$ is the sonar beam pattern, $W(r_i,t)$ is the average over a pulse length of the envelope of the pulse, and k_0 is the carrier wavenumber. The quantity $\mathbb{C}(t)$ depends on the source level, the wavenumber, propagation-loss terms, and the mechanical properties of the particles. It depends on time of reception because the propagation-loss terms depend on the round-trip length of travel of the pulse and this distance depends on time of reception—the later the time of reception, the further the pulse has propagated into the plume and the greater the propagation loss. The linear dependence on the particle volume is a result of assuming Rayleigh scattering. The intensity is given by the expression

$$I(t) = \frac{1}{c\rho} |P(t)|^2, \tag{2}$$

where c and ρ are the sound speed and density of the ocean medium, respectively, taken to be constants. If the spatial distribution of the particles is completely random, then, when averaging over the particle positions, the only terms that will survive in the double sum in Eq. (2) are the diagonal ones. The intensity is then equal to the incoherent intensity

$$I_I(t) = \frac{|\mathbb{C}(t)|^2}{c\rho} \sum_i v_i^2 D^4(\hat{r}_i) W^2(r_i, t), \tag{3}$$

we write

$$I(t) = I_I(t) + I_C(t), \tag{4}$$

where I_C is the sum of the non-diagonal terms in the double sum

$$I_C(t) = 2\frac{|\mathbb{C}(t)|^2}{c\rho}\sum_{i>j} v_i v_j D^2(\hat{r}_i)D^2(\hat{r}_j)W(r_i,t)W(r_j,t) \times \exp[2ik_0(r_i - r_j)]. \tag{5}$$

Statistical estimates of the intensity are obtained by averaging over a number of pulses while keeping the sonar orientation and operational parameters fixed so that the ensonified volume remains unchanged. These estimates are meaningful only if there is a degree of stationarity in the plume properties. While a portion, if not all, of the particles in the ensonified volume are replaced by new particles between pings because of the mass flux present, one expects that the statistical characteristics of the particles in the ensonified region and in the plume as a whole to remain unchanged in spite of this renewal process. If this were not the case then the value and meaning of the acoustical data are questionable. We assume then, that the general shape of the plume, the number of particles within it and their statistical characteristics remain unchanged over the time intervals needed to obtain a meaningful statistical average of the intensity. Consequently, the probability functions introduced below are assumed not to depend on time.

Here we employ Foldy's method for configuration averaging.[2] We assume there are N particles in the plume at a particular time. A *configuration* of the particles is given by specifying their positions (relative to the position of the transmitter) and their volumes:

$$Configuration = (\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots,\boldsymbol{r}_N|v_1,v_2,\ldots,v_N). \tag{6}$$

One is interested in average values where the average is taken over the ensemble of all possible configurations. Let $f(\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots,\boldsymbol{r}_N|v_1,v_2,\ldots,v_N)$ be some physical quantity depending on the configuration of the particles. The enemble average is then

$$\bar{f} = \int f(\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots,\boldsymbol{r}_N|v_1,v_2,\ldots,v_N) \times P(\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots,\boldsymbol{r}_N|v_1,v_2,\ldots,v_N) \times d\boldsymbol{r}_1 d\boldsymbol{r}_2,\ldots,d\boldsymbol{r}_N dv_1 dv_2,\ldots,dv_N, \tag{7}$$

where the probability distribution function $P(\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots,\boldsymbol{r}_N|v_1,v_2,\ldots,v_N)$ is defined so that

$$P(\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots,\boldsymbol{r}_N|v_1,v_2,\ldots,v_N) \times d\boldsymbol{r}_1 d\boldsymbol{r}_2,\ldots,d\boldsymbol{r}_N dv_1,dv_2,\ldots,dv_N \tag{8}$$

is the probability of finding the particles in a configuration for which the first particle lies in the element of volume $d\boldsymbol{r}_1$ centered at $\boldsymbol{r}_1$ and has a volume lying between v_1 and v_1+dv_1, the second particle lies in the element of volume $d\boldsymbol{r}_2$ centered at $\boldsymbol{r}_2$ and has a volume between v_2 and v_2+dv_2, etc.

We have

$$\int P(\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots,\boldsymbol{r}_N|v_1,v_2,\ldots,v_N) \times d\boldsymbol{r}_1 d\boldsymbol{r}_2\ldots d\boldsymbol{r}_N dv_1 dv_2\ldots dv_N = 1. \tag{9}$$

Spatial integrals and particle sums extend over all of space, i.e., all of the plume, and all of the particles in the plume. Restriction to the ensonified volume is done in a natural way because the integrals and sums include the product $D^2(\hat{r}_i)W(r_i,t)$. As we will discuss, this product defines the ensonified volume.

We consider only the situation where the probability that a particular particle is located in some volume element and has a value of v located in some range dv is independent of the locations and scattering parameters of the other particles and is the same for all particles. By making this assumption we have assumed the density of particles is small enough so that the complications that arise because of their finite volumes can be ignored. (For hydrothermal plumes the typical linear dimension of a particle is about two orders of magnitude smaller than the typical distance between particles.) We then have the factorization

$$P(\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots,\boldsymbol{r}_N|v_1,v_2,\ldots,v_N) = p(\boldsymbol{r}_1,v_1)p(\boldsymbol{r}_2,v_2)\cdots p(\boldsymbol{r}_N,v_N), \tag{10}$$

with

$$\int p(\boldsymbol{r}_i,v_i)d\boldsymbol{r}_i dv_i = 1, \quad i = 1,\ldots,N. \tag{11}$$

One can show

$$p(\boldsymbol{r},v) = \frac{n(\boldsymbol{r},v)}{N}, \tag{12}$$

where $n(\boldsymbol{r},v)dv$ is the average number of particles per unit volume at the point $\boldsymbol{r}$ having a volume between v and $v+dv$. The total number of particles per unit volume at $\boldsymbol{r}$ is

$$n(\boldsymbol{r}) = \int n(\boldsymbol{r},v)dv,$$

and the total number of particles in the plume is

$$N = \int n(\boldsymbol{r})d\boldsymbol{r}$$

The average volume of the particles at $\boldsymbol{r}$ is

$$v_{\text{ave}}(\boldsymbol{r}) = \frac{1}{n(\boldsymbol{r})}\int n(\boldsymbol{r},v)v dv. \tag{13}$$

The average of the volume squared of the particles at $\boldsymbol{r}$ is

$$v^2_{\text{rms}}(\boldsymbol{r}) = \frac{1}{n(\boldsymbol{r})}\int n(\boldsymbol{r},v)v^2 dv. \tag{14}$$

A quantity of the form of the incoherent intensity

$$\sum_i v_i^2 g(\boldsymbol{r}_i,t)$$

has the ensemble average

$$\int d\boldsymbol{r} v_{\mathrm{rms}}^2(\boldsymbol{r}) g(\boldsymbol{r},t) n(\boldsymbol{r}).$$

Whereas a quantity of the form of the coherent part of the intensity

$$2\sum_{i>j} v_i v_j g(\boldsymbol{r}_i,t) g^*(\boldsymbol{r}_j,t)$$

has the average

$$\left(1-\frac{1}{N}\right)\left|\int d\boldsymbol{r} v_{\mathrm{ave}}(\boldsymbol{r}) g(\boldsymbol{r},t) n(\boldsymbol{r})\right|^2.$$

The ensemble average of the incoherent intensity is obtained from Eq. (3):

$$\overline{I_I(t)} = \frac{|\mathbb{C}|^2}{c\rho}\int d\boldsymbol{r} v_{\mathrm{rms}}^2(\boldsymbol{r}) D^4(\hat{r}) W^2(r,t) n(\boldsymbol{r}). \tag{15}$$

For the coherent intensity, we have

$$\overline{I_C(t)} = \frac{|\mathbb{C}|^2}{c\rho}\left|\int d\boldsymbol{r} v_{\mathrm{ave}}(\boldsymbol{r}) D^2(\hat{r}) W(r,t) n(\boldsymbol{r}) e^{2ik_0 r}\right|^2,$$

where we have replaced $1-1/N$ with unity since N is assumed to always be very large.

As a point of reference, we note the coherent field $\overline{P(t)}$ obeys the relationship

$$\frac{|\mathbb{C}|^2}{c\rho}|\overline{P(t)}|^2 = \overline{I_C(t)}$$

as expected.

III. DEFINITION OF THE ENSONIFIED VOLUME

The spatial region ensonified by the sonar—the ensonified volume—is determined by the product $D^2(\hat{r})W(r,t)$. Its center is located at $\boldsymbol{r}_C = c(t-t_0)\hat{e}_{\mathrm{axis}}/2$, where $\hat{e}_{\mathrm{axis}}$ is a unit vector along the axis of the transmitter (Appendix A). Its numerical size (in units of m^2) depends on the time of reception and is given by the integral

$$\mathbb{V}(t) = \int d\boldsymbol{r} D^2(\hat{r}) W(r,t) = \int_0^\infty r^2 dr W(r,t) \int_0^\pi \sin\theta d\theta \int_0^{2\pi} d\varphi D^2(\theta,\varphi). \tag{16}$$

An average of some quantity over the ensonified volume depends not only on the size of the volume but also on its location in space. The location depends on the time, t, and the orientation of the sonar. In this paper we are keeping the orientation of the sonar fixed (the sonar is fixed relative to its platform and the platform is stationary relative to the earth) so we can suppress the dependence of averages over the ensonified volume on the orientation of the sonar and consider them only to be functions of the time of reception of the scattered wave. An integral of some function $f(\boldsymbol{r})$ over the ensonified volume is indicated by the notation

$$\int_{\mathbb{V}(t)} d\boldsymbol{r} f(\boldsymbol{r}) = \int d\boldsymbol{r} D^2(\hat{r}) W(r,t) f(\boldsymbol{r}). \tag{17}$$

Using this notation, we list

$$\overline{I_I(t)} = \frac{|\mathbb{C}|^2}{c\rho}\int_{\mathbb{V}(t)} d\boldsymbol{r} v_{\mathrm{rms}}^2(\boldsymbol{r}) D^2(\hat{r}) W(r,t) n(\boldsymbol{r}) \tag{18}$$

and

$$\overline{I_C(t)} = \frac{|\mathbb{C}|^2}{c\rho}\left|\int_{\mathbb{V}(t)} d\boldsymbol{r} v_{\mathrm{ave}}(\boldsymbol{r}) n(\boldsymbol{r}) e^{2ik_0 r}\right|^2. \tag{19}$$

We now make the assumption what while the average particle volume may be different in different parts of the plume, within a (small) ensonified volume the average and rms values for the particle volume, $v_{\mathrm{ave}}(\boldsymbol{r})$ and $v_{\mathrm{rms}}(\boldsymbol{r})$, do not change. In a sense this is the definition of what is meant by an average since the sonar cannot provide any finer spatial resolution than that of the ensonified volume. This does not mean, however, that the particle volumes themselves are all equal within the ensonified volume. We now label the average volume and the rms volume by t to indicate it only depends on the ensonified volume as a whole and move it outside the integral over the ensonified volume. We then have

$$\overline{I_I(t)} = \frac{|\mathbb{C}|^2}{c\rho} v_{\mathrm{rms}}^2(t)\int_{\mathbb{V}(t)} d\boldsymbol{r} D^2(\hat{r}) W(r,t) n(\boldsymbol{r}) \tag{20}$$

and

$$\overline{I_C(t)} = \frac{|\mathbb{C}|^2}{c\rho} v_{\mathrm{ave}}^2(t)\left|\int_{\mathbb{V}(t)} d\boldsymbol{r} n(\boldsymbol{r}) e^{2ik_0 r}\right|^2. \tag{21}$$

Consider the expression for the incoherent intensity. Regardless of how the density varies within the ensonified volume,

$$\int d\boldsymbol{r} D^4(\hat{r}) W^2(r,t) n(\boldsymbol{r})$$

is the average number of particles in the volume. Therefore, the incoherent intensity is proportional to the total number of particles in the volume, as expected. For the coherent intensity the situation is quite different. If the density is slowly varying or constant within the ensonified volume the coherent intensity is proportional to

$$\left|\int_{\mathbb{V}(t)} d\boldsymbol{r} n(\boldsymbol{r}) e^{2ik_0 r}\right|^2 \sim n^2\left|\int_{\mathbb{V}(t)} d\boldsymbol{r} e^{2ik_0 r}\right|^2$$

giving essentially zero for the reason discussed in the Introduction. On the other hand, if all the particles add coherently then the phase term in Eq. (21) is a common multiplier and we have

$$\left|\int_{\mathbb{V}(t)} d\boldsymbol{r} n(\boldsymbol{r}) e^{2ik_0 r}\right|^2 \sim n^2 \mathbb{V}^2(t),$$

the incoherent intensity is then proportional to the square of the number of particles in the ensonified volume.

IV. ADVECTION BY TURBULENT VELOCITY FLOW

The underwater plumes of interest to us are always turbulent. The particle density $n(\boldsymbol{r})$ is a passive scalar advected by the turbulent velocity flow. We assume $n(\boldsymbol{r})$ can be written as an average part and a part that fluctuates because of the turbulence

$$n(\boldsymbol{r}) = n_{\text{mean}}(\boldsymbol{r}) + \delta n(\boldsymbol{r}). \tag{22}$$

If we let $\langle \cdots \rangle$ indicate the average over the turbulent fluctuations in the plume, we have $\langle n(\boldsymbol{r}) \rangle = n_{\text{mean}}(\boldsymbol{r})$ and, for homogeneous turbulence,

$$\langle \delta n(\boldsymbol{r})\, \delta n(\boldsymbol{r}') \rangle = \delta n_{\text{rms}}^2 \int d\boldsymbol{q} \Phi(\boldsymbol{q}) \exp[i\boldsymbol{q} \cdot (\boldsymbol{r} - \boldsymbol{r}')] \tag{23}$$

where Φ is the power spectral density for the particle density, normalized so that

$$1 = \int d\boldsymbol{q} \Phi(\boldsymbol{q}). \tag{24}$$

For homogeneous, isotropic turbulence, $\Phi(\boldsymbol{q}) = \Phi(|\boldsymbol{q}|) = \Phi(q)$. (We will return to the assumptions that the turbulence is homogeneous and isotropic.) An average over the turbulent fluctuations of the incoherent intensity gives

$$\overline{\langle I_I(t) \rangle} = \frac{|\mathbb{C}|^2}{c\rho} v_{\text{rms}}^2(t) n_{\text{mean}}(t) \int_{\mathbb{V}(t)} d\boldsymbol{r} D^2(\hat{r}) W(r,t), \tag{25}$$

where we assume n_{mean} (and δn_{rms} below) depends on the ensonified volume as a whole and move it outside the integral over $\mathbb{V}$ just as with v_{rms} and v_{ave}. For the coherent intensity, we have

$$\overline{\langle I_C(t) \rangle} = \frac{|\mathbb{C}|^2}{c\rho} v_{\text{ave}}^2(t) \left(n_{\text{ave}}^2(t) \left| \int_{\mathbb{V}(t)} d\boldsymbol{r} e^{2ik_0 r} \right|^2 + \int_{\mathbb{V}(t)} d\boldsymbol{r} \int_{\mathbb{V}(t)} d\boldsymbol{r}' e^{2ik_0(r-r')} \langle \delta n(\boldsymbol{r})\, \delta n(\boldsymbol{r}') \rangle \right). \tag{26}$$

The first term can be ignored. In terms of the spectrum, we have

$$\overline{\langle I_C(t) \rangle} = \frac{|\mathbb{C}|^2}{c\rho} v_{\text{ave}}^2(t)\, \delta n_{\text{rms}}^2(t) \int d\boldsymbol{q} \Phi(q) |\mathfrak{I}(\boldsymbol{q})|^2, \tag{27}$$

where

$$\mathfrak{I}(\boldsymbol{q}) \equiv \int_{\mathbb{V}(t)} d\boldsymbol{r} \exp[ir(\boldsymbol{q} \cdot \hat{r} + 2k_0)]. \tag{28}$$

We have the expected result that the incoherent intensity depends on the average particle density and the coherent intensity depends on the covariance of the particle density. In Appendix B we show that, for a high-Q sonar, $|\mathfrak{I}(\boldsymbol{q})|^2$ is sharply peaked at the Bragg wave number $q = 2k_0$. This is not surprising because if $q \neq 2k_0$, $\mathfrak{I}(\boldsymbol{q})$ would be an integral of a phase term over many cycles of its variation. We write therefore

$$\begin{aligned} \overline{\langle I_C(t) \rangle} &\approx \frac{|\mathbb{C}|^2}{c\rho} v_{\text{ave}}^2(t)\, \delta n_{\text{rms}}^2(t) \Phi(2k_0) \int d\boldsymbol{q} |\mathfrak{I}(\boldsymbol{q})|^2 \\ &= (2\pi)^3 \frac{|\mathbb{C}|^2}{c\rho} v_{\text{ave}}^2(t)\, \delta n_{\text{rms}}^2(t) \Phi(2k_0) \\ &\quad \times \int_{\mathbb{V}(t)} d\boldsymbol{r} D^2(\hat{r}) W(r,t). \end{aligned} \tag{29}$$

This expression, which relates the temporal variation of the coherent intensity at the carrier frequency to the spatial variation of the turbulent spectrum at the Bragg wave number, $2k_0$, is usually called the *Bragg scattering condition*. In laboratory studies it has been the basis for constructing the turbulent spectrum from acoustical data for a laminar thermal plume[29] as well as for turbulent and transitional thermal plumes.[30] Such inversions require conducting measurements at different Bragg wave numbers, i.e., at different scattering angles, something that is not easily done in the ocean.

The total, averaged intensity can be written in terms of the ratio of the coherent intensity to the incoherent one, Γ,

$$\overline{\langle I(t) \rangle} = \overline{\langle I_I(t) \rangle} + \overline{\langle I_C(t) \rangle} = \overline{\langle I_I(t) \rangle}[1 + \Gamma], \tag{30}$$

where

$$\Gamma = (2\pi)^3 \left(\frac{\delta n_{\text{rms}}(t)}{n_{\text{mean}}(t)} \frac{v_{\text{ave}}(t)}{v_{\text{rms}}(t)} \right)^2 n_{\text{mean}}(t) \Phi(2k_0). \tag{31}$$

We have the remarkable result that for a high-Q sonar and within the framework of our approximations, the ratio of the coherent intensity to the incoherent one is completely independent of the characteristics of the sonar and depends only on the characteristics of the plume.[31]

If we are willing to make a few general assumptions, which are likely to be valid for many ocean plumes, we can develop this expression even further. The outer scale, L, in the turbulent flow corresponds to the size of the largest coherent structures (eddies) in the flow and for plumes this size corresponds to the linear dimension of the area through which the plume material is injected into the ocean. As an example, for a black smoker hydrothermal plume it would correspond to the diameter of the vent orifice. Since the Bragg wavelength, λ_B, for a typical plume-imaging sonar is of the order of 1 cm, the outer scale will likely be greater than the Bragg wavelength by at least an order of magnitude. In most geophysical turbulent flows, the Reynolds number, Re, will be large enough so that the inner scale, l, obtained from the relation $l = L/(\text{Re})^{3/4}$, will be very small in comparison to the Bragg wavelength. Therefore, the Bragg wavenumber is likely to be well within the inertial subrange.

Even though the geometry of a typical ocean plume does not suggest the turbulence should be isotropic, the cascade process from large scales to smaller ones within the inertial subrange is predominantly isotropic (for a discussion see Bradshaw[32]). In obtaining the estimates we seek, we can assume this local isotropy assumption is valid even though it is known not to be strictly correct (see, e.g., Ref. 33). In the same vein, we will assume the transport by the mean flow is so much greater than the transport by the Reynolds stresses that the turbulence can be considered homogeneous. From

these assumptions follow the use of a simple Kolmogorov universal spectrum, $\Phi(q) \sim 1/q^{11/3}$ in the region of the Bragg wavenumber.

For a model spectrum, we use a modification of the one developed by Duda and co-workers that has been found to be useful in several different applications.[34–36] Specifically,

$$\Phi(q) = \mathcal{N} \frac{q_L^{2/3}}{4\pi(q^2+q_L^2)^{11/6}} \frac{q^2}{q^2+q_L^2}, \quad q_l \geq q \geq q_L, \tag{32}$$

where $q_L = 2\pi/L$ and $q_l = (\mathrm{Re})^{3/4}\, q_L$ and $\mathcal{N}$ is a normalization factor close to unity. (A good approximation is $\mathcal{N} = 1.016 + 1.55/\sqrt{\mathrm{Re}}$.) In calculating this normalization factor, we have assumed the integral over q in Eq. (24) extends from q_L to q_l since the physical spectrum must drop significantly outside of this wavenumber range. A stated advantage of the model of Duda and co-workers is that it matches the Batchelor spectrum in the viscous-convective subrange where $q > q_l$ and has a diffusive roll-off. Since we are assuming the Bragg wavenumber is much less than the inner-scale wavenumber, these considerations are not significant for us. We have

$$\Phi(k_B) = \mathcal{N} \frac{(q_L/k_B)^{2/3}}{4\pi(1+(q_L/k_B)^2)^{17/6}} \frac{1}{k_B^3}, \tag{33}$$

where, again, $k_B = 2k_0$ is the Bragg wavenumber. Collecting expressions

$$\Gamma = 2\pi^2 \mathcal{N} \left(\frac{\delta n_{\mathrm{rms}}(t)}{n_{\mathrm{mean}}(t)} \frac{v_{\mathrm{ave}}(t)}{v_{\mathrm{rms}}(t)} \right)^2 \frac{(q_L/k_B)^{2/3}}{(1+(q_L/k_B)^2)^{17/6}} \frac{n_{\mathrm{mean}}(t)}{k_B^3}. \tag{34}$$

V. BLACK SMOKER HYDROTHERMAL PLUMES

As an application of Eq. (34), we consider acoustical imaging of black smoker hydrothermal plumes. Immediately above the vent orifice there is a transition region of height of about 0.25–0.5 m where both laminar and turbulent flows exist. Beyond this region the turbulence is fully developed. The outer scale, L, corresponds to the diameter of the vent orifice. According to Converse *et al.*[25] orifice diameters range from 2 to 8 cm. The outer scale wavenumber, q_L therefore varies from 0.79 to 3.1 cm^{-1}. Little *et al.*[37] estimated the Reynolds number in this region of the plume to be about 4×10^4. For the inner scale, we have $q_l = 2200–8800\ \mathrm{cm}^{-1}$. For typical plume-imaging sonars having carrier frequencies from 150 to 300 kHz, the Bragg wavenumber varies from 1.6 to 25.1 cm^{-1}. Therefore, the Bragg wavenumber is well within the inertial subrange (see Fig. 1). This would be the case even if the Reynolds number were a factor of 10 smaller. With this value for the inner scale, the normalization factor, $\mathcal{N}$, is 1.02.

The dynamic processes that take place within the plume are so rapid and the velocities so large, we can think of no mechanism that would lead to a particle size distribution dependent on location within the plume. The particles are almost surely well mixed until the plume has aged and risen hundreds of meters to near the depth of neutral buoyancy. At that stage the particle volume distribution might change because of differential settling of the larger (heavier) particles. We shall assume, therefore, that the fraction of particles having volumes between v and $v+dv$ is the same throughout the turbulent plume, i.e., $n(\boldsymbol{r},v)dv/n(\boldsymbol{r})$ is independent of $\boldsymbol{r}$. It follows v_{ave} and v_{rms} do not depend on the location of the volume ensonified by the sonar.

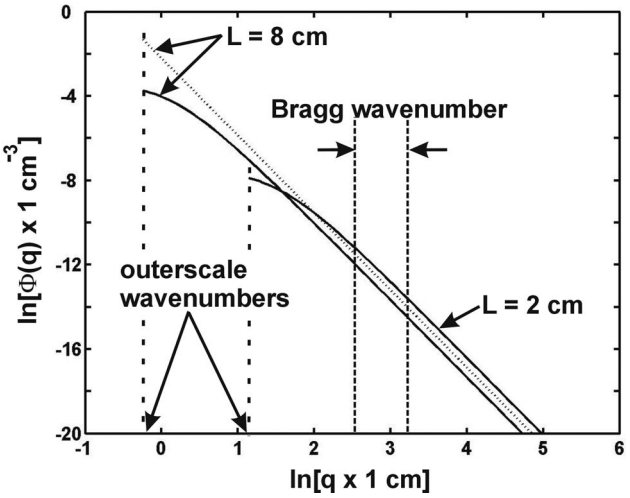


FIG. 1. Model spectra versus wave number. The solid lines are plots of the Duda and co-workers turbulent spectrum model for outer scales of 2 and 8 cm. The thin curved dashed line is the spectrum $\Phi(q) \propto 1/q^{11/3}$ for an outer scale of 8 cm. The vertical dashed lines indicate the range of Bragg wavenumbers for a typical high-frequency imaging sonar. This plot illustrates the Bragg wavenumber is well within the inertial subrange and that there is very little difference between the Duda model and the simple $1/q^{11/3}$ model at the Bragg wavenumber.

As discussed in that Ref. 19 the particles in the black smoker plumes at 21°N on the East Pacific Rise possess a bimodal size distribution, a major fraction has radii less than 5μ and a minor fraction has radii between 15μ and 35μ. We take 15μ for the average radius, which is also the value taken by Converse *et al.*[25] With the simple flat distribution used in Ref. 19, we find $v_{\mathrm{ave}}/v_{\mathrm{rms}} \sim 0.63$.

Close to the orifice Palmer and Rona[19] estimated $n_{\mathrm{mean}} \sim 30\ \mathrm{particles/cm}^3$. The concentration is, of course, reduced as the plume ages and rises in the water column. List[38] parametrized the mean values of fluid properties and species concentrations using experimental data and self-similarity arguments for submerged turbulent jets, turbulent plumes, and forced plumes and found near the axis of the plume and beyond the initial transition region, which for a black smoker plume corresponds is less than about 0.5 m, that $\delta n_{\mathrm{rms}}/n_{\mathrm{mean}} \sim 0.15–0.29$ depending on specifics.

For the range of values for the various parameters discussed above we have, using Eq. (34),[39]

$$\Gamma \sim 3.36 \times 10^{-5} – 3.40 \times 10^{-3}. \tag{35}$$

We find the coherent component is less than 1% of the incoherent component and can be ignored. The implication of this is that it is safe to assume in remote sensing applications that the backscattered intensity is proportional to the particle density multiplied by the volume being ensonified by the sonar.

VI. DISCUSSION

In this paper, we have developed an expression for the ratio of the coherent intensity to incoherent intensity for scattering from a plume of particles in the ocean assuming the coherent intensity is the result of Bragg scattering from turbulent fluctuations in the particle density. For some remote sensing applications, it is necessary that the coherent intensity be insignificant. While we use a realistic model for a high-Q pulsed sonar, it turns out the ratio does not depend on the characteristics of the sonar but only on those of the particles in the plume and those of the turbulence. As an application, we consider scattering from deep-ocean, black smoker hydrothermal plumes. We chose values for the parameters involved that characterize the central core of the plume where the coherent intensity is likely to be as large as possible. While all the values we take for the parameters can be questioned, it would be difficult to pick realistic values that would lead to a significant coherent component. The fact the coherent component is small perhaps is not surprising. The only particles that contribute to the coherent component are those that have a certain spatial relationship to one another, determined by the Bragg wavelength. Since the Bragg wavelength lies in the region where the spectrum has the $\Phi(q) \sim 1/q^{11/3}$ shape, there is nothing special or preferred about the Bragg wavelength. Consequently, only a small fraction of particles contribute to the coherent intensity.

This work assumes the scattering is in the Rayleigh scattering region where the wavelength of the acoustic field is much larger than the dimensions of the scatterer. The analysis can be generalized to other types of scattering, however. Suppose in Eq. (A13), the scattering amplitude is replaced by

$$\Lambda_f(\beta_i^{(1)}, \beta_i^{(2)}, \ldots, \beta_i^{(k)}, \cos\Theta_i)$$

where the set $[\beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(k)}]$ characterizes the scattering properties of the particles. Then a configuration consists of specifying the positions and the values of the β's for each of the particles

Configuration

$$= (\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N | \beta_1^{(1)}, \beta_2^{(1)}, \ldots, \beta_N^{(1)} | \beta_1^{(2)}, \beta_2^{(2)}, \ldots, \beta_N^{(2)} | \ldots).$$

Ensemble averaging of the incoherent and coherent intensities gives

$$\overline{I_I(t)} \propto \int_{\mathbb{V}(t)} d\boldsymbol{r} \Lambda_{\mathrm{rms}}^2(\boldsymbol{r}) D^2(\hat{r}) W(r,t) n(\boldsymbol{r})$$

and

$$\overline{I_C(t)} \propto \left| \int_{\mathbb{V}(t)} d\boldsymbol{r} \Lambda_{\mathrm{ave}}(\boldsymbol{r}) n(\boldsymbol{r}) e^{2ik_0 r} \right|^2,$$

where

$$\Lambda_{\mathrm{rms}}^2(\boldsymbol{r}) = \frac{1}{n(\boldsymbol{r})} \int n(\boldsymbol{r}, \beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(k)}) \times \Lambda_{f_0}^2(\beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(k)}, -1) d\beta^{(1)} d\beta^{(2)} \ldots d\beta^{(k)},$$

$$\Lambda_{\mathrm{ave}}(\boldsymbol{r}) = \frac{1}{n(\boldsymbol{r})} \int n(\boldsymbol{r}, \beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(k)}) \times \Lambda_{f_0}(\beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(k)}, -1) d\beta^{(1)} d\beta^{(2)} \ldots d\beta^{(k)},$$

where $n(\boldsymbol{r}, \beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(k)})$ is the obvious generalization of $n(\boldsymbol{r}, v)$. We would then end up with $v_{\mathrm{ave}}/v_{\mathrm{rms}}$ in Eq. (34) replaced with $\Lambda_{\mathrm{ave}}/\Lambda_{\mathrm{rms}}$.

In this paper, we are assuming the mean particle density, n_{mean}, is slowly varying over distances that characterize the size of the ensonified volume. This seems to be a reasonable assumption. It is worth noting, however, that if it is not valid, the mean density cannot be estimated using acoustics. On the other hand, the fluctuation in density, δn_{rms}, has spatial scales ranging from the outer scale to the inner scale. If this identification of spatial scales is valid, the mean density only contributes to the incoherent intensity and the fluctuation only to the coherent intensity. The implication of the coherent intensity, due to turbulence, being much smaller than the incoherent one is that turbulence can be completely ignored when modeling the backscattering. One needs only to consider the mean density. This result has only been demonstrated for single scattering, however. It is not clear that one can ignore the turbulence when calculating multiple scattering effects.

ACKNOWLEDGMENTS

We are grateful to Dr. C. Jones and Dr. P. A. Rona, and particularly Dr. D. R. Jackson for discussions. Dr. Jackson actually suggested that turbulence in the plume may result in a coherent component in the backscattered intensity and did preliminary calculations using a different approach than used here. We also thank Dr. Jackson and Dr. T. F. Duda for their reading of the manuscript and many valuable suggestions. We are grateful to Dr. O. A. Godin for help with the Russian references.

APPENDIX A: BACKSCATTTERED INTENSITY

In this appendix, we develop the expression for the pressure wave backscattered from a plume of particles in the ocean using a high-Q, pulsed sonar. Our approach is to model the transmitter in terms of a source term in the wave equation. Its vibrating surface is then represented by restricting the source term to a compact planar surface using a delta function.

We consider first the *transmitted wave*. Let $p^{\mathrm{trans}}(\boldsymbol{x}, t)$ be the pressure wave that propagates from the sonar transmitter. In a homogeneous, unbounded ocean without absorption p^{trans} is a solution to the equation

$$\left[\nabla^2 - \frac{1}{c^2}\partial_t^2\right] p^{\mathrm{trans}}(\boldsymbol{x}, t) = -S(t) J(\boldsymbol{x}). \tag{A1}$$

Here c is the (constant) speed of sound, $S(t)$ characterizes the temporal properties of the transmitted wave, and $J(\boldsymbol{x})$ characterizes its spatial properties. The solution to Eq. (A1) is

$$p^{\text{trans}}(\boldsymbol{x},t) = \frac{1}{4\pi}\int d\boldsymbol{y}\frac{J(\boldsymbol{y})}{|\boldsymbol{x}-\boldsymbol{y}|}S\left(t-\frac{1}{c}|\boldsymbol{x}-\boldsymbol{y}|\right). \tag{A2}$$

It is convenient to indicate locations referenced to $\boldsymbol{x}_S$, the position of the center of the active element of the transmitter. That is, we write $p^{\text{trans}}(\boldsymbol{r},t)$ rather than $p^{\text{trans}}(\boldsymbol{x},t)$ here $\boldsymbol{r}=\boldsymbol{x}-\boldsymbol{x}_S$. If $\boldsymbol{\eta}$ is the vector from the center of the transmitter to the active differential element $d\boldsymbol{\eta}$, in the far field, where $r \equiv |\boldsymbol{x}-\boldsymbol{x}_S| \gg |\boldsymbol{\eta}|$, we have

$$p^{\text{trans}}(\boldsymbol{r},t) = \frac{1}{4\pi r}\int d\boldsymbol{\eta}J(\boldsymbol{\eta})S\left(t_{\text{ret}}+\frac{\boldsymbol{\eta}\cdot\hat{r}}{c}\right), \tag{A3}$$

where $\hat{r}=\boldsymbol{r}/r$ and $t_{\text{ret}}=t-r/c$ is the retarded time.

We consider a coordinate system fixed in the active element of the transmitter and having unit vectors $\hat{e}_{\text{axis}}$, $\hat{e}_{\text{top}}$, and $\hat{e}_{\text{lat}}$, where $\hat{e}_{\text{axis}}$ is directed along the axis of the beam of the transmitter. We write $\boldsymbol{\eta}=\eta_a\hat{e}_{\text{axis}}+\boldsymbol{\eta}_\perp$, where $\boldsymbol{\eta}_\perp$ is the two-dimensional vector $\boldsymbol{\eta}_\perp=\eta_t\hat{e}_{\text{top}}+\eta_l\hat{e}_{\text{lat}}$. For a transmitter containing a planar active element, the only type we will consider, we have

$$J(\boldsymbol{\eta}) = \delta(\eta_a)j(\boldsymbol{\eta}_\perp). \tag{A4}$$

We take the function j to be an even, non-negative function of $\boldsymbol{\eta}_\perp$ normalized so that

$$\int d\boldsymbol{\eta}_\perp j(\boldsymbol{\eta}_\perp) = \int d\eta_t \int d\eta_l j(\boldsymbol{\eta}_\perp) = 1$$

A narrow-band cw pulse corresponds to

$$S(t) = E(t)\sin[2\pi f_o(t-t_0)], \tag{A5}$$

where f_o is the carrier frequency and t_0 is the time a given pulse is radiated. The slowly-varying envelope function E vanishes for $t<t_0$ and $t>t_0+\tau$. The Q of the pulse is defined by the relation $Q=f_o\tau$. Typically $Q\gg 1$. It is useful in what follows to normalize both S and E by root-mean-square values, $\tilde{S}(t)=S(t)/S_{\text{rms}}$ and $\tilde{E}(t)=E(t)/E_{\text{rms}}$. Summarizing

$$p^{\text{trans}}(\boldsymbol{r},t) = \frac{S_{\text{rms}}}{4\pi r}\int d\boldsymbol{\eta}_\perp j(\boldsymbol{\eta}_\perp)\tilde{S}\left(t_{\text{ret}}+\frac{\boldsymbol{\eta}_\perp\cdot\hat{r}}{c}\right). \tag{A6}$$

The Fourier transform of the transmitted pressure and the function $\tilde{S}$ are defined by the expressions

$$p^{\text{trans}}(\boldsymbol{r},t) = \int_{-\infty}^{\infty} p_f^{\text{trans}}(\boldsymbol{r})\exp[-2\pi i f(t-t_0)]df \tag{A7}$$

and

$$\hat{S}(t) = \int_{-\infty}^{\infty}\hat{S}_f\exp[-2\pi i f(t-t_0)]df. \tag{A8}$$

Sound absorption by the medium is described using a frequency-dependent attenuation coefficient $\alpha(f)$. It can be introduced through the replacement

$$p_f^{\text{trans}}(\boldsymbol{r}) \to 10^{-\alpha(f)r/20}p_f^{\text{trans}}(\boldsymbol{r}).$$

The Fourier transform of p^{trans} can be written as

$$p_f^{\text{trans}}(\boldsymbol{r}) = S_{\text{rms}}10^{-\alpha r/20}\frac{\exp(ikr)}{4\pi r}\tilde{S}_f D_f(\hat{r}), \tag{A9}$$

where $k\equiv 2\pi f/c$ and D_f is the beam pattern:

$$D_f(\hat{r}) \equiv \int d\boldsymbol{\eta}_\perp j(\boldsymbol{\eta}_\perp)\exp(-ik\boldsymbol{\eta}_\perp\cdot\hat{r}). \tag{A10}$$

We consider next the *scattered wave*. At the point and time of recording, defined by the position vector $\boldsymbol{r}$ and t, the total pressure wave, p^{tot}, can be written in the form

$$p^{\text{tot}}(\boldsymbol{r},t) = p^{\text{trans}}(\boldsymbol{r},t) + \sum_i p_i^{\text{scat}}(\boldsymbol{r},t). \tag{A11}$$

The contribution $p_i^{\text{scat}}(\boldsymbol{r},t)$ is the scattered pressure wave that propagates from the ith particle to the reception point at $\boldsymbol{\xi}$, arriving at time t and the sum is over all the particles. The contribution p^{trans} is the pressure wave that propagates directly from the sonar transmitter to the point of recording.

The scattered wave p_i^{scat} depends not only on the location of the recording point and the time of arrival but also on the ith particle's location, volume v_i, specific bulk e, and specific density h. The values for e and h are common to all the particles in the type of plume we are considering. We assume the ith particle can be considered stationary at $\boldsymbol{r}_i$ during the time of passage of the pulse.

The Fourier transform of p_i^{scat} is defined by

$$p_i^{\text{scat}}(\boldsymbol{r},t) = \int_{-\infty}^{\infty} p_i^{\text{scat}}(\boldsymbol{r},f)\exp[-2\pi i f(t-t_0)]df. \tag{A12}$$

The scattered wave is linearly related to the transmitted wave, evaluated at the position of the particle. We have

$$p_i^{\text{scat}}(\boldsymbol{r},f) = p_f^{\text{trans}}(\boldsymbol{r}_i)\Lambda_f(v_i,\cos\Theta_i)\frac{\exp(ik|\boldsymbol{r}-\boldsymbol{r}_i|)}{|\boldsymbol{r}-\boldsymbol{r}_i|}10^{-|\boldsymbol{r}-\boldsymbol{r}_i|\alpha(f)/20}, \tag{A13}$$

where $p_f^{\text{trans}}(\boldsymbol{r}_i)$ is given by Eq. (A9) with $\boldsymbol{r}$ replaced with $\boldsymbol{r}_i$ (i.e., $\boldsymbol{r}=\boldsymbol{x}-\boldsymbol{x}_S$ replaced by $\boldsymbol{r}_i=\boldsymbol{x}_i-\boldsymbol{x}_S$). Equation (A13) actually defines the scattering amplitude $\Lambda_f(v_i,\Theta_i)$. The quantity Θ_i is the scattering angle, given by the expression $\cos\Theta_i = \hat{r}_i\cdot(\boldsymbol{r}-\boldsymbol{r}_i)/|\boldsymbol{r}-\boldsymbol{r}_i|$. In writing Eq. (A13), we have accounted for attenuation of the pressure wave as it propagates from the ith particle to the point of recording.

For Rayleigh scattering, we have

$$\Lambda_f(v,\cos\Theta) = -\frac{k^2 v}{4\pi}\left[\frac{c-1}{c}-\frac{3(h-1)}{2h+1}\cos\Theta\right]. \tag{A14}$$

We now consider the *received wave*. The reception point is given by $\boldsymbol{r}=\boldsymbol{d}+\boldsymbol{\chi}$, where $\boldsymbol{d}$ is the position vector of the center of active element of the receiver and $\boldsymbol{\chi}$ is a vector to some infinitesimal portion of the active element of the receiver. It is analogous to $\boldsymbol{\eta}$. The pressure waves sensed by the infinitesimal active elements of the receiver are coherently added at any instant to form the received pressure wave

$$p^{\mathrm{rec}}(t) = \int d\boldsymbol{\chi} J'(\boldsymbol{\chi}) p^{\mathrm{tot}}(\boldsymbol{d}+\boldsymbol{\chi},t). \tag{A15}$$

The function J' describes how the infinitesimal active elements are weighted in forming the received pressure. We introduce an additional coordinate system with its center at $\boldsymbol{d}=0$ and having unit vectors $\hat{e}'_{\mathrm{axis}}$, $\hat{e}'_{\mathrm{top}}$, $\hat{e}'_{\mathrm{lat}}$ such that the active element of the receiver is along $\hat{e}'_{\mathrm{axis}}$, with $\hat{e}'_{\mathrm{axis}}$ directed *into* the ocean. Therefore,

$$J'(\boldsymbol{\chi}) = \delta(\chi_a) j'(\boldsymbol{\chi}_\perp), \tag{A16}$$

where $\boldsymbol{\chi}=\chi_a \hat{e}'_{\mathrm{axis}}+\boldsymbol{\chi}_\perp$ with $\boldsymbol{\chi}_\perp=\chi_t \hat{e}'_{\mathrm{top}}+\chi_l \hat{e}'_{\mathrm{lat}}$. We take the function j' to depend on $\boldsymbol{\chi}_\perp$ in the same way as j to depends on $\boldsymbol{\eta}_\perp$. The direct arrival p^{trans} is eliminated from p^{tot} in a well-designed experiment. So

$$p^{\mathrm{rec}}(t) = \int d\boldsymbol{\chi}_\perp j'(\boldsymbol{\chi}_\perp) \sum_i p_i^{\mathrm{scat}}(\boldsymbol{d}+\boldsymbol{\chi}_\perp,t).$$

The Fourier transform is

$$p_f^{\mathrm{rec}} = \int d\boldsymbol{\chi}_\perp j'(\boldsymbol{\chi}_\perp) \sum_i p_i^{\mathrm{scat}}(\boldsymbol{d}+\boldsymbol{\chi}_\perp,f).$$

With $\boldsymbol{R}_i \equiv \boldsymbol{r}_i - \boldsymbol{d}$, where $|\boldsymbol{\chi}_\perp| \ll R_i$ for all $\boldsymbol{\chi}_\perp$ of interest, we have

$$p_f^{\mathrm{rec}} = \sum_i p_f^{\mathrm{trans}}(\boldsymbol{x}_i) \Lambda_f(v_i,\cos\Theta_i) 10^{-R_i\alpha(f)/20} \frac{\exp(ikR_i)}{R_i} D'_f(\hat{R}_i),$$

where D'_f is the beam pattern of the receiver

$$D'_f(\hat{R}_i) = \int d\boldsymbol{\chi}_\perp j'(\boldsymbol{\chi}_\perp) \exp(-ik\boldsymbol{\chi}_\perp \cdot \hat{R}_i).$$

Collecting expressions gives

$$p_f^{\mathrm{rec}} = \frac{S_{\mathrm{rms}} \hat{S}_f}{4\pi} \sum_i \Lambda_f(v_i,\cos\Theta_i) 10^{-\alpha(f)\bar{r}_i/10} \times \frac{\exp(2ik\bar{r}_i)}{r_i R_i} D_f(\hat{r}_i) D'_f(\hat{R}_i), \tag{A17}$$

where $\bar{r}_i=(r_i+R_i)/2$ and $\cos\Theta_i=-\hat{r}_i\cdot\hat{R}_i$.

We assume (as is customary) several of the quantities in Eq. (A17) are so slowly varying across the bandwidth that they can be approximated by their values at the carrier frequency. They are the scattering amplitude, i.e., the k^2 multiplier, the attenuation coefficient, and the beam patterns.

$$\Lambda_f(v,\cos\Theta) \approx \Lambda_{f_0}(v,\cos\Theta) \equiv \Lambda(v,\cos\Theta),$$

$$\alpha(f) \approx \alpha(f_0) \equiv \alpha,$$

$$D_f(\hat{r}_i) \approx D_{f_0}(\hat{r}_i) \equiv D(\hat{r}_i), \quad D'_f(\hat{R}_i) \approx D'_{f_0}(\hat{R}_i) \equiv D'(\hat{R}_i).$$

In the time domain, we then have

$$p^{\mathrm{rec}}(t) = \frac{S_{\mathrm{rms}}}{4\pi} \sum_i D(\hat{r}_i) \frac{10^{-\alpha\bar{r}_i/10}}{r_i R_i} \Lambda(v_i,\cos\Theta_i) D'(\hat{R}_i) \times \tilde{S}\left(t - t_0 - \frac{2\bar{r}_i}{c}\right).$$

For the narrow-band, cw pulse we are considering we can construct the quadrature components in the usual way and from them the complex demodulated pressure

$$\mathfrak{p}^{\mathrm{rec}} = \sqrt{2} c^{i\pi/2} \frac{S_{\mathrm{rms}}}{4\pi} \sum_i D(\hat{r}_i) \Lambda(v_i,\cos\Theta_i) \frac{10^{-\alpha\bar{r}_i/10}}{r_i R_i} \times D'(\hat{R}_i) \hat{E}\left(t - t_0 - \frac{2\bar{r}_i}{c}\right) \exp(2ik_0\bar{r}_i). \tag{A18}$$

This procedure is equivalent to stating that the physical pressure is the real part of a quantity involving the complex demodulated pressure, i.e., $p^{\mathrm{rec}}=\mathrm{Re}[p^{\mathrm{rec}} \exp(-2\pi i f_0(t-t_0))]$.

If the wave scattered by the ith particle is to significantly contribute to the pressure recorded at time t, it must be within the sonar's ensonified volume at an earlier time, t less the time it takes for the wave to travel from the particle to the receiver. The location and shape of the ensonified volume depend on the orientation and characteristics of the active elements of the sonar transmitter and receiver as well as the time of reception. We let t_C be the time it takes for the transmitted wave to reach the center of the ensonified volume. The time interval t_C will depend on t. For example, a monostatic sonar has a center located at $\vec{r}_C(t)=(1/2)c(t-t_0)\hat{e}_{\mathrm{axis}}$ relative to the sonar transmitter and the time it takes for the transmitted wave to reach the center of the ensonified volume is $t_C(t)=(t-t_0)/2$. If the ith particle is to contribute to the pressure recorded at time t, it must be within the ensonified volume as it existed at time $t_C(t)$. For narrow-beam pulsed sonars, the linear dimensions of the ensonified volume are always small compared to the distances from the sonar elements to the center of the ensonified volume. That is, with $\boldsymbol{R}_C=\boldsymbol{r}_C-\boldsymbol{d}$, we have $|\boldsymbol{r}_i-\boldsymbol{r}_C| \ll r_C$ and $|\boldsymbol{R}_i-\boldsymbol{R}_C| \ll R_C$. This allows us to replace r_i and R_i in the spreading and absorption loss terms with r_C and R_C, respectively, and $\cos\Theta_i$ with $\cos\Theta_C=-\hat{r}_C\cdot\hat{R}_C$. Equation (A22) becomes

$$\mathfrak{p}^{\mathrm{rec}}(t) = \mathbb{C}(t) \sum_i v_i D(\hat{r}_i) D'(\hat{R}_i) \hat{E}\left(t - \frac{2\bar{r}_i}{c}\right) \exp(2ik_0\bar{r}_i), \tag{A19}$$

where we have introduced

$$\mathbb{C}(t) \equiv \sqrt{2} e^{-i\pi/2} S_{\mathrm{rms}} \left(\frac{10^{-\alpha\bar{r}_C/20} k_0}{4\pi\sqrt{r_C R_C}}\right)^2 \left(\frac{e-1}{e} + \frac{3(h-1)}{2h+1} \cos\Theta_C\right) \tag{A20}$$

and $\vec{r}_C \equiv (r_C+R_C)/2$ to simplify the notation. This coefficient depends on the time of recording, t, since r_C and R_C depend on this time.

Finally, we perform a running average over an interval of time equal to the pulse width and starting at time $t-t_0$,

$$P(t) = \frac{1}{\tau} \int_{t-t_0}^{t-t_0-\tau} \mathfrak{p}^{\mathrm{rec}}(t')dt' = \mathbb{C}(t) \sum_i v_i D(\hat{r}_i) D'(\hat{R}_i) W(\bar{r}_i,t) \exp(2ik_0\bar{r}_i) \tag{A21}$$

where

$$W(r,t)=\frac{1}{\tau}\int_{t-t_0}^{t-t_0+\tau}\tilde{E}\left(t'-\frac{2r}{c}\right)dt' \tag{A22}$$

defines a range gate.

We now restrict the discussion to a monstatic sonar, i.e., one for which the transmitter serves also as the receiver. For such a sonar, $R_C=r_C$, $\boldsymbol{r}_i=\boldsymbol{R}_i$ so $\bar{r}_i=r_i$. Also $\Theta_C=\pi$, and $D'(\hat{R}_i)=D(\hat{r}_i)$. The pressure is then given by Eq. (1) with

$$\mathbb{C}(t)=\sqrt{2}e^{-i\pi/2}S_{\mathrm{rms}}\left(\frac{10^{-\alpha r_C(t)/20}k_0}{4\pi r_C(t)}\right)^2\left(\frac{e-1}{e}-\frac{3(h-1)}{2h+1}\right). \tag{A23}$$

APPENDIX B: THE INTEGRAL $\mathfrak{I}$

Equation (28) can be written as

$$\mathfrak{I}=\int_0^\infty r^2 dr W(r,t)\exp(2irk_0)T_1(r\boldsymbol{q}), \tag{B1}$$

with

$$T_1(r\boldsymbol{q})\equiv\int_0^\pi \sin\theta d\theta\int_0^{2\pi}d\varphi D^2(\hat{r})\exp(ir\boldsymbol{q}\cdot\hat{r}). \tag{B2}$$

Let L be the characteristic dimension of the active element of the transducer (not to be confused with the outer scale of turbulence in the body of the paper). So $j(\boldsymbol{\eta}_\perp)=0$ for $|\boldsymbol{\eta}_\perp|>L$. Let $\boldsymbol{\beta}_\perp=\boldsymbol{\eta}_\perp/L$ then

$$T_1(r\boldsymbol{q})=L^4\int d\boldsymbol{\beta}_\perp j(L\boldsymbol{\beta}_\perp)\int d\boldsymbol{\beta}'_\perp j(L\boldsymbol{\beta}'_\perp)\int_0^\pi \sin\theta d\theta \times\int_0^{2\pi}d\varphi\exp[ik_0(r\boldsymbol{q}/k_0-L(\boldsymbol{\beta}_\perp+\boldsymbol{\beta}'_\perp))\cdot\hat{r}] \tag{B3}$$

Using the well-known relation

$$\int_0^\pi \sin\theta d\theta\int_0^{2\pi}d\varphi\exp(i\boldsymbol{a}\cdot\hat{r})=4\pi\frac{\sin(a)}{a},$$

we have

$$T_1(r\boldsymbol{q})=4\pi L^4\int d\boldsymbol{\beta}_\perp j(L\boldsymbol{\beta}_\perp)\int d\boldsymbol{\beta}'_\perp j(L\boldsymbol{\beta}'_\perp)\times\frac{\sin(k_0|r\boldsymbol{q}/k_0-L(\boldsymbol{\beta}_\perp+\boldsymbol{\beta}'_\perp)|)}{k_0|r\boldsymbol{q}/k_0-L(\boldsymbol{\beta}_\perp+\boldsymbol{\beta}'_\perp)|}.$$

In this expression, $|\boldsymbol{q}/k_0|$ and $|\boldsymbol{\beta}_\perp+\boldsymbol{\beta}_\perp|$ are of order unity and $r\gg L$. Therefore, as is usually done in other contexts, we replace the denominator with rq and keep terms linear in L in the argument of the sine function

$$\frac{\sin(k_0|r\boldsymbol{q}/k_0-L(\boldsymbol{\beta}_\perp+\boldsymbol{\beta}'_\perp)|)}{k_0|r\boldsymbol{q}/k_0-L(\boldsymbol{\beta}_\perp+\boldsymbol{\beta}'_\perp)|}\approx\frac{1}{rq}\sin(rq-Lk_0\hat{q}\cdot(\boldsymbol{\beta}_\perp+\boldsymbol{\beta}'_\perp))$$

so, approximately,

$$T_1(r\boldsymbol{q})=\frac{4\pi}{rq}\int d\boldsymbol{\eta}_\perp j(\boldsymbol{\eta}_\perp)\times\int d\boldsymbol{\eta}'_\perp j(\boldsymbol{\eta}'_\perp)\sin(rq-k_0\hat{q}\cdot(\boldsymbol{\eta}_\perp+\boldsymbol{\eta}'_\perp)).$$

Using the symmetry of the j-functions, we can write this as

$$T_1(r\boldsymbol{q})=4\pi\frac{\sin(rq)}{rq}\int d\boldsymbol{\eta}_\perp j(\boldsymbol{\eta}_\perp)\times\int d\boldsymbol{\eta}'_\perp j(\boldsymbol{\eta}'_\perp)\exp[-ik_0\hat{q}\cdot(\boldsymbol{\eta}_\perp+\boldsymbol{\eta}'_\perp)]=4\pi\frac{\sin(rq)}{rq}D^2(\hat{q}). \tag{B4}$$

We have

$$\mathfrak{I}(\boldsymbol{q})=\frac{4\pi}{q}D_0^2(\hat{q})T_2, \tag{B5}$$

with

$$T_2=\int_0^\infty rdrW(r,t)\exp(2irk_0)\sin(rq). \tag{B6}$$

With a change in variable $r=c\tau((t-t_0)/\tau+x)/2$, we have

$$T_2=\left(\frac{c\tau}{2}\right)^2\exp(ik_0c(t-t_0))\times\int_{-1}^1\left(\frac{t-t_0}{\tau}+x\right)dxW\left(\frac{c\tau}{2}\left(\frac{t-t_0}{\tau}+x\right),t\right)\times\exp(ik_0c\tau x)\sin\left(\frac{cq}{2}(t-t_0+\tau x)\right). \tag{B7}$$

Since $t-t_0\gg\tau$, this can be approximated with

$$T_2=\left(\frac{c\tau}{2}\right)^2\left(\frac{t-t_0}{\tau}\right)W\left(\frac{c}{2}(t-t_0),t\right)\exp(ik_0c(t))\times\int_{-1}^1 dx\exp(ik_0c\tau x)\sin\left(\frac{cq}{2}(t-t_0+\tau x)\right)=-\frac{1}{2i}\left(\frac{c\tau}{2}\right)^2\left(\frac{t-t_0}{\tau}\right)\times W\left(\frac{c(t-t_0)}{2},t\right)\times\exp\left[2\pi icf_0(t-t_0)\left(1-\frac{q}{2k_0}\right)\right]\times\frac{\sin 2\pi Q(1-q/2k_0)}{\pi Q(1-q/2k_0)}. \tag{B8}$$

A term proportional to

$$\frac{\sin 2\pi Q(1+q/2k_0)}{\pi Q(1+q/2k_0)}$$

has been ignored since it makes a negligible contribution because q and k_0 are non-negative. Collecting expressions gives

$$|\Im(\boldsymbol{q})|^2 = \left(\frac{4\pi(t-t_0)}{q\tau}\right)^2 \left(\frac{c\tau}{2} D(\hat{q})\right)^4 W^2\left(\frac{c}{2}(t-t_0), t\right) \times \left(\frac{\sin 2\pi Q(1-q/2k_0)}{2\pi Q(1-q/2k_0)}\right)^2. \quad \text{(B9)}$$

Since $Q \gg 1$, this expression is sharply peaked at the Bragg wavenumber $q=2k_0$ as advertised.

[1] Lord Rayleigh, *Theory of Sound*, (Dover, New York, 1945), Vol. **1**, Sec. 42a.
[2] L. L. Foldy, "The multiple scattering of waves," Phys. Rev. **67**, 107–119 (1945).
[3] M. Lax, "Multiple scattering of waves," Rev. Mod. Phys. **23**, 287–310 (1951).
[4] M. Lax, "Multiple scattering of waves II," Phys. Rev. **85**, 621–629 (1952).
[5] P. C. Waterman and R. Truell, "Mutiple scattering of waves," J. Math. Phys. **2**, 512–537 (1961).
[6] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (IEEE, Piscataway, NJ, 1997).
[7] P. M. Morse and K. U. Ingard, *Theoretical Acoustics*, (McGraw-Hill, New York, 1968), pp. 436–441.
[8] J. C. Grace, "Reflection from a statistical array of scatterers," J. Acoust. Soc. Am. **26**, 103–104 (1954).
[9] V. P. Glotov, "Coherent scattering of plane and spherical waves in deep-sea layers containing discrete inhomogeneities," Dokl. Akad. Nauk SSSR **143**, 312–315 (1961) [Sov. Phys. Dokl. **7**, 211–213 (1962)].
[10] V. P. Glotov, "Coherent scattering of sound from clusters of discrete inhomogeneities in pulsed transmission," Akust. Zh. **8**, 281–284 (1962) [Sov. Phys. Acoust. **8**, 220–222 (1963)].
[11] V. P. Glotov and Yu. P. Lysanov, "The scattered field for a spherical source above a plane layer containing discrete inhomogeneities," Akust. Zh. **9**, 176–181 (1963) [Sov. Phys. Acoust. **9**, 142–146 (1963)].
[12] B. F. Kur'yanov, "Coherent and Incoherent scattering of waves by a set of point scatterers distributed randomly in space," Akust. Zh. **10**, 195–201 (1963) [Sov. Phys. Acoust. **10**, 160–164 (1964)].
[13] See, for example, S. Machlup and J. B. Hersey, "Analysis of sound-scattering observations from non-uniform distributions of scatterers in the ocean," Deep-Sea Res. **3**, 1–22 (1955).
[14] T. G. Leighton, *The Acoustic Bubble* (Academic, New York, 1994).
[15] Our use of the terms "coherent" and "incoherent" is consistent with their use in the underwater acoustics literature. It should be noted, however, in the literature on Bragg volume scattering, the scattered field is usually referred to as "incoherent." We thank D. R. Jackson for pointing out this semantic inconsistency.
[16] For a review, see D. R. Palmer, "Acoustical imaging of naturally-occurring underwater plumes," in *Acoustical Imaging*, edited by J. Jones (Plenum, New York, 1995), Vol. **21**.
[17] See, for example, D. R. Palmer, "On the interpretation of measurements of acoustic backscatter from dredged-material plumes," J. Marine Env. Engg. **7**, 125–152 (2003).
[18] D. R. Palmer and P. A. Rona, "Acoustical imaging of deep ocean hydrothermal flows," in *Sounds in the Sea from Ocean Acoustics to Acoustical Oceanography*, edited by H. Medwin (Cambridge University Press, Cambridge, 2005).
[19] D. R. Palmer, P. A. Rona, and M. J. Mottl, "Acoustic imaging of high-temperature hydrothermal plumes at seafloor spreading centers," J. Acoust. Soc. Am. **80**, 888–898 (1986).
[20] P. A. Rona, D. R. Jackson, K. G. Bemis, C. D. Jones, K. Milsuzawa, D. R. Palmer, and D. Silver, "Acoustics advances study of sea floor hydrothermal flow," EOS Trans. Am. Geophys. Union **83**, 497–502 (2002).
[21] K. G. Bemis, P. A. Rona, D. R. Jackson, C. D. Jones, D. Silver, and K. Mitsuzawa, "A comparison of black smoker hydrothermal plume behavior at Monolith Vent and at Clam Acres vent field: Dependence on source configuration," Mar. Geophys. Res. **23**, 81–96 (2002).
[22] P. A. Rona, K. G. Bemis, D. Silver, and C. D. Jones, "Acoustic imaging, visualization, and quantification of buoyant hydrothermal plumes in the ocean," Mar. Geophys. Res. **23**, 147–168 (2002).
[23] B. R. Morton, G. I. Taylor, and J. S. Turner, "Turbulent gravitational convection from maintained and instantaneous sources," Proc. R. Soc. London, Ser. A **234**, 1–23 (1956).
[24] See, for example, P. D. Thorne, P. J. Hardcastle, and R. L. Soulsby, "Analysis of acoustic measurements of suspended sediments," J. Geophys. Res. **98**, 899–910 (1993).
[25] D. R. Converse, H. D. Holland, and J. M. Edmond, "Flow rates in the axial hot springs of the East Pacific Rise (21°N): Implications for the heat budget and the formation of massive suflide deposits," Earth Planet. Sci. Lett. **69**, 159–175 (1984).
[26] D. R. Palmer, "Rayleigh scattering from nonspherical particles," J. Acoust. Soc. Am. **99**, 1901–1912 (1996).
[27] D. R. Palmer, "Acoustical scattering from constituents of an ocean plume located near a boundary surface," IEEE Trans. Geosci. Remote Sens. **43**, 770–777 (2005).
[28] As an aside, there is a clear need for numerical simulations of scattering from plumes in the ocean to guide experimental design and the interpretation of data. This material could be the first step in developing algorithms for doing these simulations.
[29] L. Goodman, J. Oeschger, and D. Szargowicz, "Ocean acoustics turbulence study: Acoustic scattering from a buoyant axisymmetric plume," J. Acoust. Soc. Am. **91**, 3212–3227 (1992).
[30] J. Oeschger and L. Goodman, "Acoustic scattering from a thermally driven buoyant plume revisited," J. Acoust. Soc. Am. **113**, 1353–1367 (2003).
[31] D. R. Jackson pointed out that this result may not be so remarkable after all. The incoherent to coherent intensity ratio, Γ, may essentially be the ratio of volume scattering strengths and one would expect these volume scattering strengths to be independent of sonar properties. We have only shown this independence for a high-Q sonar.
[32] P. Bradshaw, "Chapter 1: Introduction," in *Turbulence*, edited by P. Bradshaw (Springer-Verlag, New York, 1978).
[33] R. H. Kraichnan, "On Kolmogorov's inertial-range theories," J. Fluid Mech. **62**, 305–330 (1974).
[34] T. F. Duda, S. M. Flatté, and D. B. Creamer, "Modelling meter-scale acoustic intensity fluctuations from oceanic fine structure and microstructure," J. Geophys. Res. **93**, 5130–5142 (1988).
[35] T. F. Duda, "Modeling weak fluctuations of undersea telemetry signals," Inf. Sci. (N.Y.) **16**, 3–11 (1991).
[36] T. F. Duda, "Predicting scattering of sound by diffuse hydrothermal vent plumes at mid-ocean ridges," J. Acoust. Soc. Am. **103**, 330–335 (1998).
[37] S. A. Little, K. D. Stolzenbach, and R. P. Von Herzen, "Measurements of plume flow from a hydrothermal vent field," J. Geophys. Res. **92**, 2587–2596 (1987).
[38] E. J. List, "Mechanics of turbulent buoyant jets and plumes," in *Turbulent Buoyant Jets and Plumes*, edited by W. Rodi (Pergamon, New York, 1982).
[39] T. F. Duda pointed out that if, indeed, the Bragg wavenumber is in the inertial subrange, we could have equally well used the model spectrum $\Phi(q) \propto 1/q^{11/3}$. Had we done this, $\mathcal{N}/(1+(q_L/k_B)^2)^{17/6}$ in Eq. (34) would have been replaced with unity. For typical values this would have led to a change in the numerical values listed for Γ, Eq. (35), by about 9%; an amount that does not alter any of our conclusions.

Deep seafloor arrivals: An unexplained set of arrivals in long-range ocean acoustic propagation

Ralph A. Stephen and S. Thompson Bolmer
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543-1542

Matthew A. Dzieciuch and Peter F. Worcester
Scripps Institution of Oceanography, UCSD, La Jolla, California 92093-0225

Rex K. Andrew, Linda J. Buck, and James A. Mercer
University of Washington, Seattle, Washington 98105-6698

John A. Colosi
Naval Postgraduate School, Monterey, California 93943

Bruce M. Howe
University of Hawaii at Manoa, Honolulu, Hawaii 96822

(Received 5 November 2008; revised 12 May 2009; accepted 15 May 2009)

Receptions, from a ship-suspended source (in the band 50–100 Hz) to an ocean bottom seismometer (about 5000 m depth) and the deepest element on a vertical hydrophone array (about 750 m above the seafloor) that were acquired on the 2004 Long-Range Ocean Acoustic Propagation Experiment in the North Pacific Ocean, are described. The ranges varied from 50 to 3200 km. In addition to predicted ocean acoustic arrivals and deep shadow zone arrivals (leaking below turning points), "deep seafloor arrivals," that are dominant on the seafloor geophone but are absent or very weak on the hydrophone array, are observed. These deep seafloor arrivals are an unexplained set of arrivals in ocean acoustics possibly associated with seafloor interface waves.
 [DOI: 10.1121/1.3158826]

PACS number(s): 43.30.Gv, 43.30.Re, 43.30.Nb, 43.30.Ma [RCG] Pages: 599–606

I. INTRODUCTION

The Long-Range Ocean Acoustic Propagation Experiment (LOAPEX) was carried out in the North Pacific Ocean between 10 September and 10 October 2004 (Mercer *et al.*, 2005, 2009, 2006). Two goals of LOAPEX were to understand the role of bottom interaction in long-range, low-frequency acoustic propagation, and to understand the physical mechanisms responsible for the so-called "deep shadow zone arrivals" observed by Dushaw *et al.* (1999). Deep shadow zone arrivals occur when acoustic energy is scattered vertically many wavelengths below a turning point (Van Uffelen *et al.*, 2008, 2009, 2006) and examples will be shown below. The 2004 LOAPEX experiment was well suited to address these issues in three respects. First, broadband acoustic transmissions (in the band 50–100 Hz) were simultaneously received on a pair of vertical hydrophone arrays spanning 3500 m of the water column and on four ocean bottom seismometer/hydrophones (OBS/Hs). The primary receiver was a deep vertical line array (DVLA) consisting of 60 hydrophones from nominal depths of 2150–4270 m (Worcester, 2005). A shallow vertical line array was deployed about 5 km west of the DVLA and consisted of 40 hydrophones at nominal depths from 350 to 1750 m. The four OBS/Hs rested on the seafloor and were located about 2 km west, north, east, and south of the DVLA. Second, the seven ship-suspended source locations varied from 50 to 3200 km from the DVLA (labeled T50, T250, etc.) so that the evolution of the shadow zone arrivals with range could be observed. Finally, at each source station signals were transmitted over intervals from 9 to 34 h so that the acoustic sensitivity to oceanic processes with time scales on the order of minutes to hours could be addressed.

II. THE EXPERIMENT

A. Sources

The acoustic source was suspended at depths of 350, 500, or 800 m and transmitted primarily phase-coded M-sequences [short for "binary maximal-length sequences" (Munk *et al.*, 1995)] with a bandwidth from about 50 to 100 Hz. Each M-sequence lasted approximately 30 seconds and sequential transmissions lasted for periods of 20 to 80 min. The source parameters and transmission schedule are given in Mercer *et al.* (2009). For simplicity in this paper results are shown only for phase-coded M-sequences with a 68.2 Hz carrier frequency and a 2 cycle per digit code modulation rate (henceforth referred to as M68.2 sequences) with the source at 350 m depth for ranges from 250 to 3200 km (only six combinations of source range, source depth, and transmission format from over 20 possible permutations) (Fig. 1 and Table I). Broadband source levels in this case were about 194 dB re 1 μPa at 1 m as derived from the sound pressure level measured at a monitoring hydrophone (Mercer *et al.*, 2005).

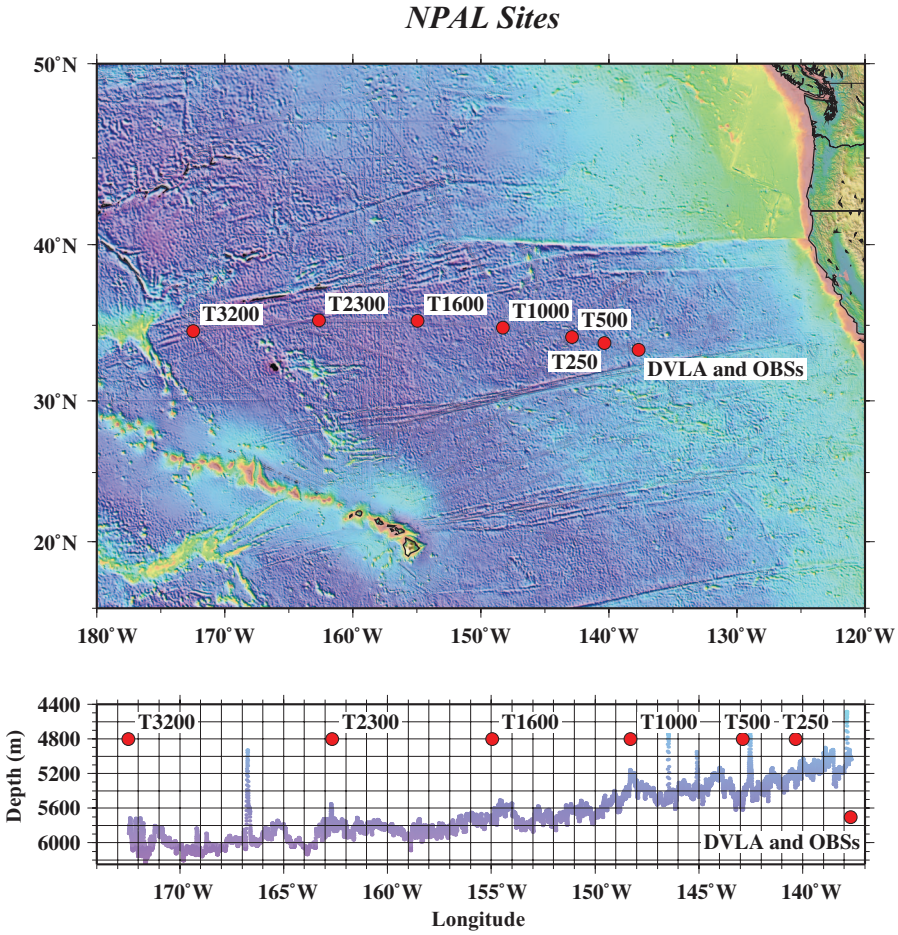


FIG. 1. The locations of the sources and receivers discussed in this paper are shown on a map of the North Pacific with the satellite-derived bathymetry (Smith and Sandwell, 1997). The geodetic lines from all of the transmission stations to the DVLA and South OBS coincide within 2 km (Fig. 2). The bathymetry along this geodetic line is shown as a function of longitude in the lower figure where the source and receiver longitudes are given as red dots. The bathymetry along this geodetic line is deeper than 4400 m everywhere.

B. Receivers

In this report a preliminary comparison of the vertical component geophone data (responding to vertical particle velocity) from the south OBS/H on the seafloor (at 4973 m depth) with data from the deepest DVLA hydrophone (4250 m depth) (only two receiver channels from over 100 available) is presented. These are labeled OBS-S-Geo and DVLA-L20-Hyd, respectively.

C. Processing

Pulse-like arrivals with improved resolution (27 ms in time, 40 m in range) and signal-to-noise ratio (SNR) were obtained by replica correlation (also called matched filtering or pulse compression, see Munk *et al.*, 1995) applied to individual received sequences. Sequences were not summed together prior to the replica correlation. The SNR was further improved by incoherently stacking the magnitude of the replica-correlated traces. The magnitude of the traces was simply summed without regarding the phase of the complex output of the correlation process. The durations of the transmissions at each station and the number of acceptable sequences that were included in the stacked traces for OBS-S-Geo and DVLA-L20-Hyd are given in Table II. A discussion of the processing, with examples, and comparisons with other analyses being carried out on the LOAPEX data set is given in Stephen *et al.* (2008).

D. Bathymetry

The locations of the sources and receivers discussed in this paper are given in Table I and are shown, overlain on

TABLE I. Source and receiver locations for the data presented in this paper.

	Latitude	Longitude
DVLA-L20-Hyd	33° 25.1′ N	137° 40.9′ W
OBS-S-Geo	33° 23.9′ N	137° 41.0′ W
T250	33° 52.2′ N	140° 19.4′ W
T500	34° 14.9′ N	142° 52.9′ W
T1000	34° 51.9′ N	148° 16.8′ W
T1600	35° 17.1′ N	154° 57.0′ W
T2300	35° 18.8′ N	162° 38.9′ W
T3200	34° 37.9′ N	172° 28.4′ W

TABLE II. Approximate elapsed times and the number of acceptable sequences (NN_OBS and NN_DVLA for OBS-S-Geo and DVLA-L20-Hyd, respectively) used for the stacked traces in Figs. 4 and 5.

	Elapsed time (h)	NN_OBS	NN_DVLA
T250	9	421	27
T500	15	690	480
T1000	34	1345	1080
T1600	28	975	930
T2300	14	606	576
T3200	15	599	576

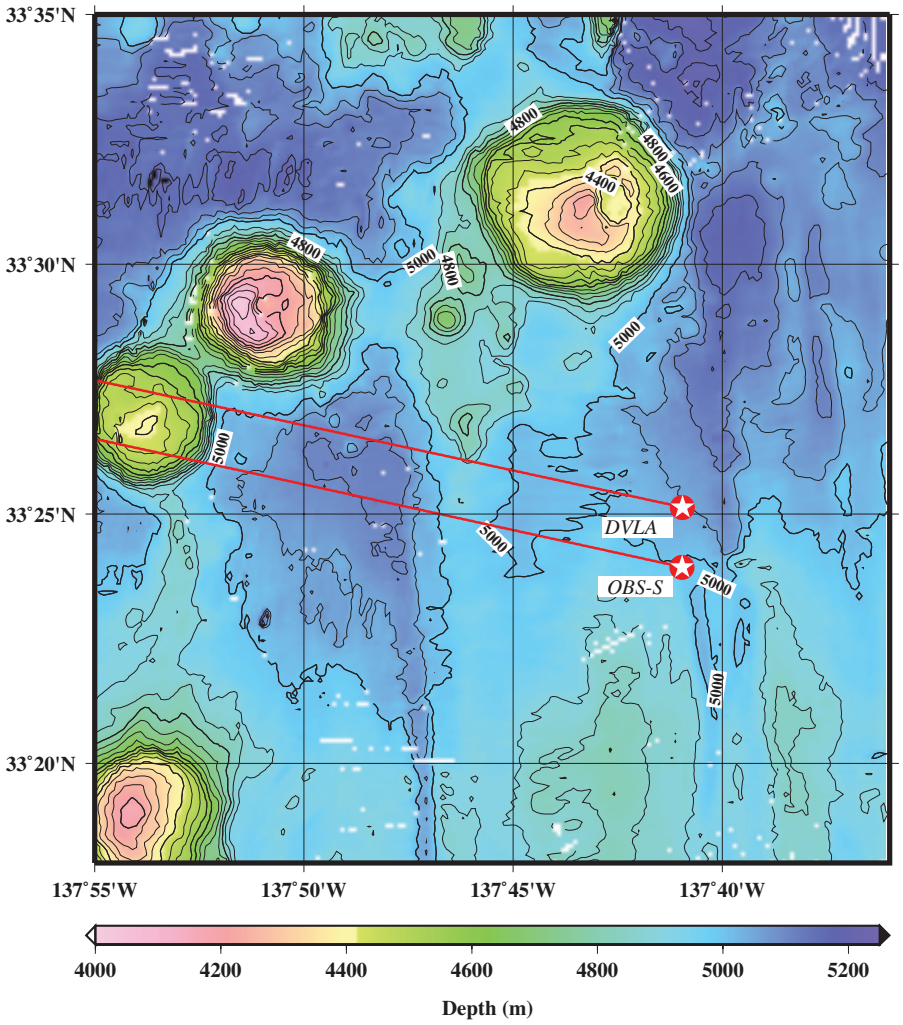


FIG. 2. The swath-mapped bathymetry within about 10 nm of the DVLA (Worcester, 2005) shows bottom features, as shallow as 4000 m, that may contribute to the arrival structure discussed in this paper. The geodetic lines to the source locations are shown as red lines. The sources were positioned to lie on the same geodetic line to the DVLA. Propagation paths to the DVLA and OBS-S are coincident within 2 km.

satellite-derived bathymetry (Smith and Sandwell, 1997), in Fig. 1. The transmitting stations were chosen so that they fall on the same geodetic line from the DVLA. The ocean depth as a function of longitude along the geodetic is given at the bottom of Fig. 1. The bathymetry is deeper than 4400 m everywhere along this geodetic line. There was a slight offset between the South OBS/H and the DVLA, but the geodetic lines to each are within 2 km (Fig. 2). Swath bathymetry was acquired during the experiments in 2004 from the DVLA out to T1000 (Worcester, 2005). At the resolution of the bathymetry in Fig. 1 there are no new features across the swath, about 2 km either side of the geodetic line to the DVLA.

Figure 2 shows the swath bathymetry within about 10 nm of the DVLA. There are four hills, all deeper than 4000 m, that could conceivably play a role, via horizontal refraction and bottom interaction, in the arrival structure at OBS-S-Geo.

E. Sound speed profiles

Figure 3 shows the sound speed profiles based on conductivity temperature-depth (CTD) casts acquired at the transmission stations during the 2004 experiment (Mercer *et al.*, 2005). The maximum and minimum bounds of the sound speed profiles from the World Ocean Atlas (Antonov *et al.*, 2006; Locarnini *et al.*, 2006), that were used for the PE modeling below, are overlain for comparison.

F. Parabolic equation (PE) modeling

To aid in the interpretation of the records, as a preliminary step, the observations are compared to PE model predictions (Collins and Westwood, 1991) based on range-dependent bathymetry from Smith and Sandwell (1997) and sound speed profiles from the 2005 World Ocean Atlas. Specifically the RAMGEO program (Collins, 1993) was used to synthesize the model records. This is a wide-angle energy-conserving Padé PE propagation model. Internal waves were not included in these models. Any internal wave scattering would generate the *Z*-waves (discussed below) (Van Uffelen *et al.*, 2009) and not the *S*-waves, so including them would not offer a likely explanation for the *S*-waves.

Initially the PE modeling consisted of two strategies. The first strategy was compressional wave modeling without bottom interaction (keep the bottom properties the same as the water above it but add strong attenuation so that no energy is returned from the seafloor or sub-seafloor). This strategy, without including bottom interaction, has successfully predicted long-range, ocean acoustic propagation in the past

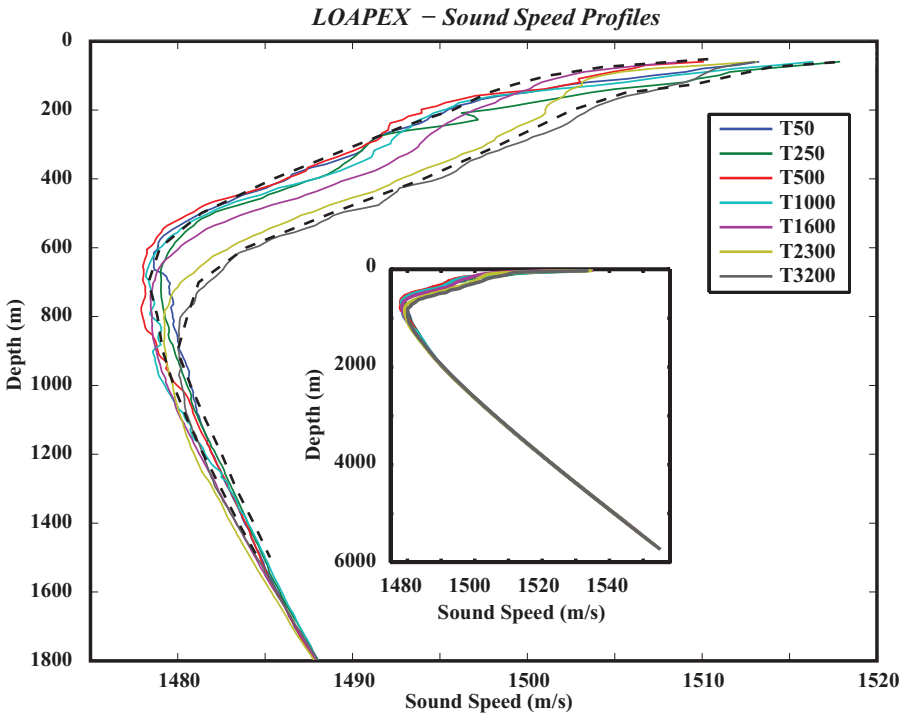


FIG. 3. The sound speed profiles that were acquired at each source location during the experiment (see Fig. 1) are shown as colored solid lines (Mercer *et al.*, 2005). The maximum and minimum sound speeds as a function of depth from the World Ocean Atlas, that were used for the PE modeling, are shown as black, dotted lines. The profiles overlap below about 1400 m.

using a variety of methods (Colosi *et al.*, 2005; de Groot-Hedlin *et al.*, 2009; Dushaw *et al.*, 1999; Heaney *et al.*, 1991; Van Uffelen *et al.*, 2009; Wage *et al.*, 2005; Xu, 2007). This seemed like a good initial strategy since bathymetry along the whole 3200 km long geodetic is everywhere deeper than 4400 m and for most of the propagation path is deeper than 5000 m (Fig. 1).

The second strategy was compressional wave modeling with bottom interaction. In this case the model is about 12 km thick. The seafloor consisted of (i) a 20 m thick layer of homogeneous sediment with $Vp=1.6$ km/s and attenuation of 0.01 dB/m at 70 Hz [all attenuation values are from Hamilton (1976), although better values for sediments have been recommended in more recent papers (Bowles, 1997; Kibblewhite, 1989; Mitchell and Focke, 1980)], (ii) a 2 km thick layer of basalt with a gradient in *P*-wave speed from 4.0 to 6.8 km/s and attenuation of 0.0025 dB/m, (iii) a 4 km thick layer of gabbro with a gradient in *P*-wave speed from 6.8 to 8.1 km/s and attenuation of 0.0025 dB/m, and (iv) a homogeneous half-space for the mantle at 8.1 km/s and attenuation of 0.0025 dB/m. Density in the sediments (mostly pelagic clay) is given by: density (g/cc)=1.35 +(1.80−1.35)/300×depth (m) (Hamilton, 1976). For the igneous rocks density is related to compressional sound speed by: density (g/cc)=1.91+0.158*Vp* (km/s) (Swift *et al.*, 1998).

Neither explicit seafloor roughness (distinct from the large-scale range-dependent bathymetry) nor shear wave properties in the bottom were included in the PE models. The point of this paper is that the arrival structure on the seafloor geophone is distinctly different from the arrival structure on a hydrophone 750 m above the seafloor. Future analysis will include modeling that considers the additional elastic waves (shear and interface waves) and scattering from seafloor roughness and sub-seafloor heterogeneity (Collins, 1989, 1991; Stephen and Swift, 1994; Swift and Stephen, 1994; Wetton and Brooke, 1990).

III. ARRIVAL CLASSES

The results of the preliminary analysis show that the arrival structure in the seafloor data has similarities and differences from the arrival structure on the DVLA. In particular, the first arrivals on the OBS/H geophone and the deepest DVLA hydrophone correspond to energy in the first deep arriving path predicted by the PE model. In addition, some of the later arrivals correspond to energy leaking from shallower turning points above the receiver; these are the so-called shadow zone arrivals previously described. Importantly, some of the later arrivals on the geophone record are not observed in the DVLA or model records. These "deep seafloor" arrivals therefore do not correspond to any previously recognized oceanic propagation path. These signals are, however, often the largest events observed on the deep seafloor at long ranges (up to 3200 km).

All of the model results and data in both Figs. 4 and 5, described below, correspond to M68.2 sequence transmissions at 350 m depth. The data traces in both figures were computed by incoherent summing of all acceptable replica-correlated sequences. Since the plotted traces are normalized to the maximum amplitude on the trace, the "sum" and the "average" plot the same. The number of "acceptable" sequences differed between the hydrophone and geophone channels because of different recording windows and noisy or spiky traces that were excluded from the sums. The number of "good" sequences and the total elapsed time at each station are summarized in Table II.

Some caveats of this preliminary analysis are as follows. First, all of the replica correlations presented in this paper

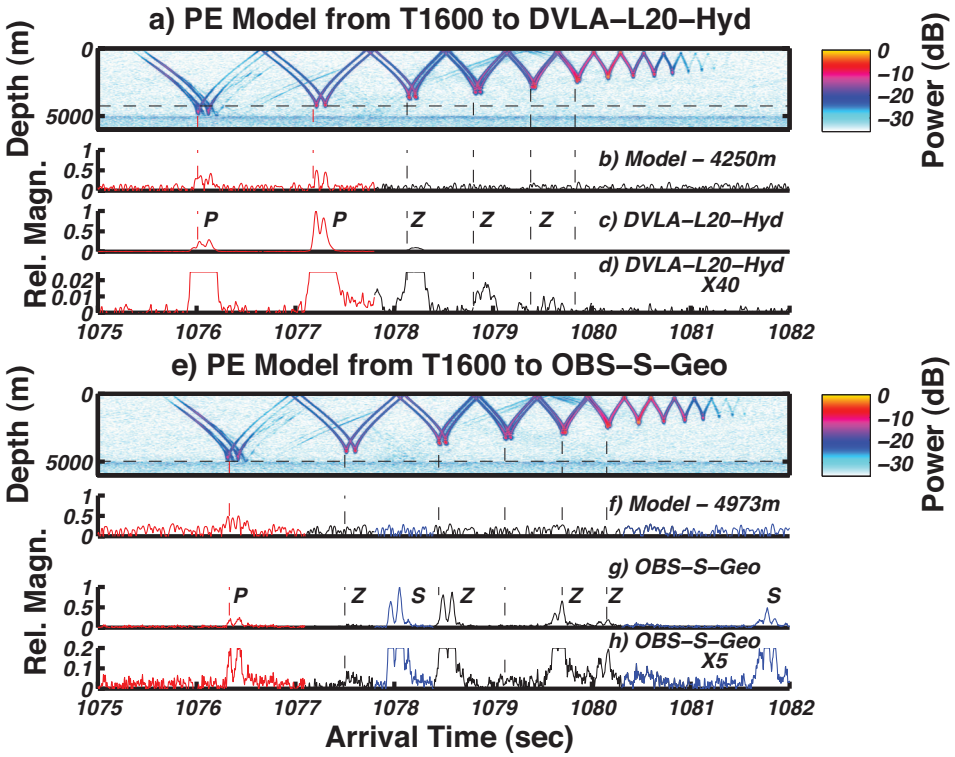


FIG. 4. This figure compares the arrival structure on DVLA-L20-Hyd and OBS-S-Geo with PE model predictions for a range of 1600 km. The PE models include bottom interaction. The top group of four panels (a)–(d) is the model-data comparison for DVLA-L20-Hyd and the bottom group [(e)–(h)] is for OBS-S-Geo. Within each group of four, the top panel is the time front diagram, the second panel is the model trace at the receiver depth (indicated by a horizontal dashed line in the time front diagram), the third panel is the data trace normalized to its maximum amplitude, and the bottom trace is an expanded view of the data trace. Vertical dashed lines show the times of the turning points across all of the plots. Examples of the three arrival classes, "PE predicted" arrivals (P), "deep shadow zone" arrivals (Z), and "deep seafloor" arrivals (S) are indicated. The deep seafloor arrivals are an unexplained set of arrivals.

were computed assuming that the sources and receivers were stationary—no corrections for motion of the sources and receivers have been made. A preliminary analysis of Doppler effects on the data presented here (Stephen *et al.*, 2008) indicates that this is a valid assumption. Furthermore the geophones and hydrophones on the OBS/Hs were both self-noise limited so that only upper bounds can be placed on the true seafloor ambient noise, and the SNRs are minimum values (Stephen *et al.*, 2006).

A. DVLA-L20-Hyd arrivals

Figure 4(a) shows the predicted time front using the PE method for an M68.2 sequence transmission at 350 m depth to a range of about 1600 km from the DVLA. This calculation includes bottom interaction. The time front between about 1080.5 and 1081.3 s corresponds to refracted-refracted (RR) paths, for which energy stays trapped in the sound channel and propagates over very long ranges with very little

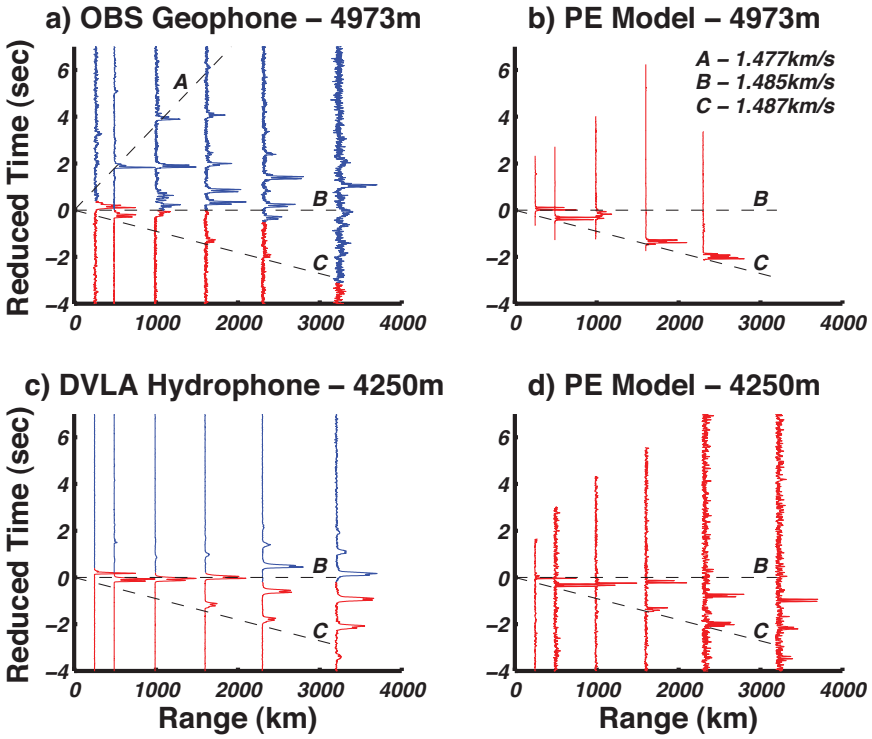


FIG. 5. The stacked traces from the OBS vertical geophone on the seafloor (a) show many more arrivals than the deepest DVLA hydrophone (c) or the PE models [(b) and (d)]. For the OBS geophone traces (a), events occurring with a sound speed faster than about 1.485 km/s (roughly earlier than line *B*) are predicted by the PE but there are many "late arrivals." Dashed lines correspond to three relevant speeds: (A) the apparent sound speed of the latest arrival at T500, T1000, and T1600, (B) the apparent sound speed of the largest PE arrivals at the deepest hydrophone of the DVLA which seems to separate the known early arrivals from the late unknown arrivals, and (C) the apparent sound speed of the earliest arriving energy at the OBS and DVLA, which corresponds to the deepest turning energy (see Fig. 4). The time axis has been reduced by subtracting the range divided by 1.485 km/s.

attenuation (Munk *et al.*, 1995). The time front from about 1075 to 1080.5 s corresponds to refracted surface-reflected (RSR) paths. At these frequencies there is little scattering loss on reflection from the ocean surface and energy on RSR paths also propagates to long ranges with little attenuation. (Sea states throughout the experiment were generally calm with the roughest conditions, up to sea state 3, occurring between T1600 and T2300.) The lower turning points (or caustics) of the RR and RSR paths form a progressively shallower sequence as time increases. The RSR paths (near 1076.1 and 1076.2 s, for example) are typically brightest near the lower turning points because of focusing at the caustic. At the lower turning points, which occur above the seafloor even for the earliest arriving event, the grazing angle is zero.

Time fronts corresponding to weak surface-reflected bottom-reflected (SRBR) paths, which often attenuate very rapidly with range and are often unnecessary to successfully predict long-range, deep-water, propagation (de Groot-Hedlin *et al.*, 2009; Heaney *et al.*, 1991; Van Uffelen *et al.*, 2009; Wage *et al.*, 2003), can be seen as more horizontal time fronts at mid-water depths between 1076.5 and 1077.5 s and at shallow depths between 1078 and 1080 s.

The arrival structure at DVLA-L20-Hyd [Figs. 4(c) and 4(d)] corresponds well with the modeled time fronts [Figs. 4(a) and 4(b)] for the earlier travel times. The first two lower turning points in the time front (at about 1076 and 1077 s) reach the hydrophone depth [indicated by the horizontal dashed line in Fig. 4(a)] and clear arrivals are observed at these times. For convenience the arrivals that match the time fronts are referred to as "PE predicted" arrivals. In this example, the magnitude of the PE predicted arrivals increases with subsequent turning points.

The next two arrivals on DVLA-L20-Hyd (at about 1078 and 1078.7 s) occur at times corresponding to predicted turning points above the receiver depth. These "deep shadow zone" arrivals, which occur at about the same time as shallower turning points in the time fronts (Dushaw *et al.*, 1999), can be attributed to diffraction and scattering by internal waves (leakage) below the turning points (Van Uffelen *et al.*, 2009, 2006). The magnitude of the deep shadow zone arrivals decreases with subsequent turning points, as expected for decay below the progressively shallower turning point depths. There is even a weak indication of a third deep shadow zone arrival at about 1079.4 s.

B. OBS-S-Geo arrivals

The arrival structure on OBS-S-Geo [Figs. 4(g) and 4(h)] is very different from DVLA-L20-Hyd [Figs. 4(c) and 4(d)] or the PE model. There are more arrivals and their spacing is less regular. The first, weak doublet on the geophone trace occurs below the deepest and earliest turning point in the time front. This is a PE predicted arrival.

Of the four large amplitude later arrivals, only the arrivals near 1078.4 and 1079.5 s appear to correspond to a turning point and could be called deep shadow zone arrivals. The large magnitude arrivals near 1078.0 and 1081.7 s do not correspond to turning points. These are an unexplained set of arrivals called deep seafloor arrivals. The first deep seafloor arrival, occurring near 1078.0 s and the largest event on OBS-S-Geo, occurs between two prominent arrivals that could be called deep shadow zone arrivals because they are beneath adjacent turning points (the weak arrival near 1077.5 s and the strong arrival near 1078.4 s). The second deep seafloor arrival at about 1081.7 s is occurring after the finale time. That these arrivals are not a result of leakage from turning points of the time front is supported by the fact that these arrivals are not observed on DVLA-L20-Hyd. Further these arrivals do not correspond to the PE predicted SRBR time fronts [Figs. 4(e) and 4(f)].

The arrival at about 1078.4 s is the second largest event on OBS-S-Geo and it coincides with the third turning point (from the left). The arrival at about 1079.5 s is the third largest event on OBS-S-Geo and it coincides with the fifth turning point. These are labeled as deep shadow zone arrivals since they coincide with turning points in the PE model and there are corresponding events on DVLA-L20-Hyd (although the event below the fifth turning point is very weak). The intensity pattern on the OBS-S-Geo record is then quite curious because the deep shadow zone arrivals do not become progressively weaker as travel time increases. Deep shadow zone arrivals typically get much weaker with subsequent, shallow turning points (as on DVLA-L20-Hyd, for example). This is not the case for OBS-S-Geo and these could be deep seafloor arrivals.

IV. RECORD SECTIONS AND PROPAGATION SPEEDS

In Fig. 5, record sections of the stacked traces for OBS-S-Geo are compared to similar sections for DVLA-L20-Hyd and to predictions based on the PE model. The PE model results in Fig. 5(b) do not include bottom interaction because the results for the OBS with bottom interaction, for example, Fig. 4(f), were too noisy to show meaningful arrival structure. The PE model results for 4250 m depth [Fig. 5(d)] include simple bottom interaction as in Fig. 4(a). For simplicity, just the transmissions from the LOAPEX source at 350 m depth are considered.

A comparison of Fig. 5(a) with Fig. 5(c) readily shows that OBS-S-Geo has a very different arrival structure than DVLA-L20-Hyd at ranges from 500 to 3200 km. Up to 2300 km range the first geophone arrival corresponds to the first arrival on the deepest DVLA hydrophone. Comparisons of Fig. 5(a) with Fig. 5(b) (up to 2300 km range) and of Fig. 5(c) with Fig. 5(d) (up to 3200 km range) show that the earliest arrivals on OBS-S-Geo and on DVLA-L20-Hyd are kinematically predicted by the PE model and fall on or near the propagation sound speed of 1.487 km/s, line *C*.

The late, large amplitude arrivals on OBS-S-Geo mostly occur after a slower propagation speed of 1.485 km/s [line *B* on Fig. 5(a)]. As discussed above and shown in Figs. 4(e)–4(h), many of these large amplitude arrivals do not correspond to turning points. Many of the events that do occur at turning point times have amplitudes that are inconsistent with purely waterborne deep shadow zone arrivals. The late arrivals on OBS-S-Geo are up to 20 dB larger than the earliest arrivals.

On OBS-S-Geo [Fig. 5(a)] at 500 and 1000 km range, there are large amplitude arrivals occurring after the finale time and with an apparent speed of 1.477 km/s (line *A*) that is slower than even the slowest sound channel minimum measured on the geodetic (about 1.478 km/s, Fig. 3). There is even a weak event at 1600 km range corresponding to this apparent speed. Neither of these large amplitude events is observed on the DVLA hydrophone or the PE models. These are clear examples of deep seafloor arrivals.

V. SUMMARY

Receptions of stacked, replica-correlated traces on a single OBS geophone on the seafloor are compared with a hydrophone moored about 750 m above the seafloor. The OBS geophone generally has more arrivals than the moored hydrophone. Two types of arrivals on the OBS geophone are observed that are not explained by PE modeling using simple sound speed profiles. The "deep shadow zone" arrivals occur at the time of shallower turning points (Dushaw *et al.*, 1999; Van Uffelen *et al.*, 2009), are consistent with decay from shallower turning points, are also observed on the DVLA hydrophones, and their arrival time is predicted by PE propagation models. The deep seafloor arrivals, on the other hand, occur later than the first PE arrival, are not readily observed on the DVLA hydrophones, and their arrival time is not predicted by PE propagation models. There are even strong arrivals after the PE predicted finale region. Deep seafloor arrivals are among the largest events observed at the seafloor. This is an unexplained set of arrivals in long-range ocean acoustic propagation.

The observed intensity pattern of the OBS arrivals is significantly more complex than the waterborne arrivals seen on the DVLA. The deep shadow zone arrivals observed on the OBS and associated with shallower turning points do not display the expected decay of intensity as travel time increases. This observation could be due to a few factors. First, the acoustic energy reaching the OBS will naturally have more bottom interaction, thus modulating the intensity pattern. Second, there could, in fact, be some interference between deep shadow zone arrivals and deep seafloor arrivals. Third, the association of some OBS arrivals with turning points may be a coincidence—all of the OBS arrivals could be deep seafloor arrivals. Fourth, in some instances the arrivals on the OBS could be associated with SRBR paths. Further analysis will be required to resolve these issues.

Deep seafloor arrivals appear to be an interface wave whose amplitude decays upward into the water column. The interface wave could be a shear-related mode coupled to the sound channel propagation (Butler, 2006; Butler and Lomnitz, 2002; Park *et al.*, 2001) or it could be excited by secondary scattering from bottom features (Chapman and Marrett, 2006; Dougherty and Stephen, 1988; Schreiner and Dorman, 1990). These unexplained arrivals could conceivably be horizontal multi-path from some persistent ocean thermal structure, but it would be necessary to explain why they are observed on the seafloor OBS but not on the DVLA only 2 km away.

In this letter the existence of deep seafloor arrivals has been simply addressed and a little of their kinematics (arrival times) as observed on the geophone of one OBS has been compared with one DVLA hydrophone. In a later paper the results from the two other OBS geophones, from the OBS hydrophones, and from other hydrophones in the DVLA will be presented and discussed. The propagation physics of the arrivals will also be discussed through a quantitative analysis of signal amplitudes, ambient and system noise, and SNRs.

ACKNOWLEDGMENTS

The idea to deploy OBSs on the 2004 NPAL experiment was conceived at a workshop held at Woods Hole in March 2004 (Odom and Stephen, 2004). The LOAPEX source deployments, the moored DVLA receiver deployments, and some post-cruise data reduction and analysis were funded by the Office of Naval Research under Award Nos. N00014-1403-1-0181, N00014-03-1-0182, and N00014-06-1-0222. Additional post-cruise analysis support was provided to RAS through the Edward W. and Betty J. Scripps Chair for Excellence in Oceanography. The OBS/Hs used in the experiment were provided by Scripps Institution of Oceanography under the U.S. National Ocean Bottom Seismic Instrumentation Pool (SIO-OBSIP—http://www.obsip.org). To cover the costs of the OBS/H deployments funds were paid to SIO-OBSIP from the National Science Foundation and from the Woods Hole Oceanographic Institution Deep Ocean Exploration Institute. The OBS/H data are archived at the IRIS (Incorporated Research Institutions for Seismology) Data Management Center.

Antonov, J. I., Locarnini, R. A., Boyer, T. P., Mishonov, A. V., and Garcia, H. E. (**2006**). *World Ocean Atlas 2005*, Vol. **2**, edited by S. Levitus (U.S. Government Printing Office, Washington, DC), NOAA Atlas NESDIS 62.

Bowles, F. A. (**1997**). "Observations on attenuation and shear-wave velocity in fine-grained, marine sediments," J. Acoust. Soc. Am. **101**, 3385–3397.

Butler, R. (**2006**). "Observations of polarized seismoacoustic T waves at and beneath the seafloor in the abyssal Pacific ocean," J. Acoust. Soc. Am. **120**, 3599–3606.

Butler, R., and Lomnitz, C. (**2002**). "Coupled seismoacoustic modes on the seafloor," Geophys. Res. Lett. **29**, 57-1-57-4.

Chapman, N. R., and Marrett, R. (**2006**). "The directionality of acoustic T-phase signals from small magnitude submarine earthquakes," J. Acoust. Soc. Am. **119**, 3669–3675.

Collins, M. D. (**1989**). "A higher-order parabolic equation for wave propagation in an ocean overlying an elastic bottom," J. Acoust. Soc. Am. **86**, 1459–1464.

Collins, M. D. (**1991**). "Higher-order Padé approximations for accurate and stable elastic parabolic equations with application to interface wave propagation," J. Acoust. Soc. Am. **89**, 1050–1057.

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

Collins, M. D., and Westwood, E. K. (**1991**). "A higher-order energy-conserving parabolic equation for range-dependent ocean depth, sound speed, and density," J. Acoust. Soc. Am. **89**, 1068–1075.

Colosi, J. A., Baggeroer, A. B., Cornuelle, B. D., Dzieciuch, M. A., Munk, W. H., Worcester, P. F., Dushaw, B. D., Howe, B. M., Mercer, J. A., Spindel, R. C., Birdsall, T. D., Metzger, K., and Forbes, A. M. G. (**2005**). "Analysis of multipath acoustic field variability and coherence in the finale of broadband basin-scale transmissions in the North Pacific Ocean," J. Acoust. Soc. Am. **117**, 1538–1564.

de Groot-Hedlin, C., Blackman, D. K., and Jenkins, C. S. (**2009**). "Effects of variability associated with the Antarctic circumpolar current on sound propagation in the ocean," Geophys. J. Int. **176**, 478–490.

Dougherty, M. E., and Stephen, R. A. (**1988**). "Seismic energy partitioning and scattering in laterally heterogeneous ocean crust," Pure Appl. Geo-

phys. **128**, 195–229.
Dushaw, B. D., Howe, B. M., Mercer, J. A., and Spindel, R. (**1999**). "Multimegameter-range acoustic data obtained by bottom-mounted hydrophone arrays for measurement of ocean temperature," IEEE J. Ocean. Eng. **24**, 202–214.
Hamilton, E. L. (**1976**). "Sound attenuation as a function of depth in the seafloor," J. Acoust. Soc. Am. **59**, 528–535.
Heaney, K. D., Kuperman, W. A., and McDonald, B. E. (**1991**). "Perth–Bermuda sound propagation (1960): Adiabatic mode interpretation," J. Acoust. Soc. Am. **90**, 2586–2594.
Kibblewhite, A. C. (**1989**). "Attenuation of sound in marine sediments: A review with emphasis on new low-frequency data," J. Acoust. Soc. Am. **86**, 716–738.
Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., and Garcia, H. E. (**2006**). *World Ocean Atlas 2005*, Vol. **1**, edited by S. Levitus (U.S. Government Printing Office, Washington, DC), NOAA Atlas NESDIS 61.
Mercer, J., Andrew, R., Howe, B. M., and Colosi, J. (**2005**). "Cruise report: Long-Range Ocean Acoustic Propagation experiment (LOAPEX)," Report No. APL-UW TR 0501, Applied Physics Laboratory, University of Washington, Seattle, WA.
Mercer, J. A., Colosi, J. A., Howe, B. M., Dzieciuch, M. A., Stephen, R., and Worcester, P. F. (**2009**). "LOAPEX: The Long-Range Ocean Acoustic Propagation Experiment," IEEE J. Ocean. Eng. **34**, 1–11.
Mercer, J. A., Howe, B. M., Andrew, R. K., Wolfson, M. A., Worcester, P. F., Dzieciuch, M. A., and Colosi, J. A. (**2006**). "The Long-Range Ocean Acoustic Propagation Experiment (LOAPEX): An overview," J. Acoust. Soc. Am. **120**, 3020.
Mitchell, S. K., and Focke, K. C. (**1980**). "New measurements of compressional wave attenuation in deep ocean sediments," J. Acoust. Soc. Am. **67**, 1582–1589.
Munk, W., Worcester, P., and Wunsch, C. (**1995**). *Ocean Acoustic Tomography* (Cambridge University Press, Cambridge, UK).
Odom, R. I., and Stephen, R. A. (**2004**). Proceedings, Seismo-Acoustic Applications in Marine Geology and Geophysics Workshop, Woods Hole Oceanographic Institution, [Report No. APL-UW TR 0406, Applied Physics Laboratory, University of Washington, Seattle, WA].
Park, M., Odom, R. I., and Soukup, D. J. (**2001**). "Modal scattering: A key to understanding oceanic T-phases," Geophys. Res. Lett. **28**, 3401–3404.
Schreiner, M. A., and Dorman, L. M. (**1990**). "Coherence lengths of seafloor noise: Effect of ocean bottom structure," J. Acoust. Soc. Am. **88**, 1503–1514.
Smith, W. H. F., and Sandwell, D. T. (**1997**). "Global seafloor topography from satellite altimetry and ship depth soundings," Science **277**, 1956–1962.
Stephen, R. A., and Swift, S. A. (**1994**). "Modeling seafloor geoacoustic interaction with a numerical scattering chamber," J. Acoust. Soc. Am. **96**, 973–990.
Stephen, R. A., Mercer, J. A., Andrew, R. K., and Colosi, J. A. (**2006**). "Seafloor hydrophone and vertical geophone observations during the North Pacific Acoustic Laboratory/Long-range Ocean Acoustic Propagation Experiment (NPAL/LOAPEX)," J. Acoust. Soc. Am. **120**, 3021.
Stephen, R. A., Bolmer, S. T., Udovydchenkov, I., Worcester, P. F., Dzieciuch, M. A., Van Uffelen, L., Mercer, J. A., Andrew, R. K., Buck, L. J., Colosi, J. A., and Howe, B. M. (**2008**). "NPAL04 OBS data analysis part 1: Kinematics of deep seafloor arrivals," Report No. WHOI-2008-03, Woods Hole Oceanographic Institution, Woods Hole, MA.
Swift, S. A., and Stephen, R. A. (**1994**). "The scattering of a low-angle pulse beam by seafloor volume heterogeneities," J. Acoust. Soc. Am. **96**, 991–1001.
Swift, S. A., Lizarralde, D., Hoskins, H., and Stephen, R. A. (**1998**). "Seismic attenuation in upper oceanic crust at Hole 504B," J. Geophys. Res. **103**, 27193–27206.
Van Uffelen, L., Worcester, P., and Dzieciuch, M. (**2008**). "Absolute intensities of acoustic shadow zone arrivals," J. Acoust. Soc. Am. **123**, 3464.
Van Uffelen, L. J., Worcester, P. F., Dzieciuch, M. A., and Rudnick, D. L. (**2009**). "The vertical structure of shadow-zone arrivals at long range in the ocean," J. Acoust. Soc. Am. **125**, 3539-3588.
Van Uffelen, L. J., Worcester, P. F., Dzieciuch, M. A., Rudnick, D. L., Cornuelle, B. D., and Munk, W. H. (**2006**). "The vertical structure of shadow-zone arrivals at long range in the ocean (A)," J. Acoust. Soc. Am. **119**, 3344.
Wage, K. E., Baggeroer, A. B., and Preisig, J. C. (**2003**). "Modal analysis of broadband acoustic receptions at 3515-km range in the North Pacific using short-time Fourier techniques," J. Acoust. Soc. Am. **113**, 801–817.
Wage, K. E., Dzieciuch, M. A., Worcester, P. F., Howe, B. M., and Mercer, J. A. (**2005**). "Mode coherence at megameter ranges in the North Pacific Ocean," J. Acoust. Soc. Am. **117**, 1565–1581.
Wetton, B. T. R., and Brooke, G. H. (**1990**). "One-way wave equations for seismoacoustic propagation in elastic waveguides," J. Acoust. Soc. Am. **87**, 624–632.
Worcester, P. (**2005**). "North Pacific Acoustic Laboratory: SPICE04 Recovery Cruise Report," Scripps Institution of Oceanography, La Jolla, CA.
Xu, J. (**2007**). "Effects of internal waves on low frequency, long range, acoustic propagation in the deep ocean," Ph.D. thesis, MIT/WHOI Joint Program, Woods Hole, MA.

Coherence function for the stochastic scattering by a time-varying, slightly rough, acoustically soft surface

Robert W. Scharstein[a)]
Department of Electrical Engineering, University of Alabama, 317 Houser Hall, Tuscaloosa, Alabama 35487-0286

Richard S. Keiffer
Naval Research Laboratory, Code 7181, Stennis Space Center, Mississippi 39529

(Received 4 February 2009; revised 23 May 2009; accepted 28 May 2009)

Approximations for the coherence function of the random acoustic field scattered by a moving rough surface are extracted from a second-order perturbation expansion. The statistically rough surface is characterized by a simple, wide-sense stationary autocorrelation function that exhibits temporal and spatial dispersive behavior in terms of three parameters: the acoustically-small mean-square surface height, a correlation length, and a group velocity. Excitation is provided by the quintessential obliquely-incident time-harmonic plane-wave. Asymptotic evaluation of Fourier integral representations of the stochastic field yields expressions for the coherence function that explicitly display the dependence upon the space and time coordinates and their interaction via the geometrical parameters of the rough surface. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3158927]

PACS number(s): 43.30.Hw, 43.30.Re [AMJD] Pages: 607–611

I. INTRODUCTION

The fundamental importance and physical interpretation of the mutual coherence function of the acoustic field scattered by the time-varying, rough surface of the ocean are summarized by Parkins,[1] who considered plane-wave excitation of a finite patch of surface. The familiar $\sin x/x$ angular behavior due to such "finite aperture" sources is prevalent in all of his results. Clay and Medwin[2] followed a similar course and focus on the coherence between partial field contributions scattered by separate subareas. Both Parkins[1] and Clay and Medwin[2] used a first-order correction to the Kirchhoff–Helmholtz integral. Harper and Labianca[3] made serious analytic progress by applying perturbation theory to two different time-varying rough surfaces; one having a spatially uniform autocorrelation function and one comprised of a finite sum of deterministic, sinusoidal surface waves. The results of Harper and Labianca[3] explicitly display the expected Doppler effects and additional complications due to an inhomogeneous (refractive) medium. Acoustically large surface roughness is treated by Dowling and Jackson[4] using quadrature on the surface integral in the Kirchhoff or physical acoustics approximation. More in line with the philosophy of the present paper is the work of Shaw *et al.*[5] who sought explicit, analytic expressions for the coherence functions of the scattered field by starting with reasonably simple, yet physically meaningful, rough surface correlation functions. They derived asymptotic expansions for the Kirchhoff integral. Unlike the present paper, their scattering surfaces do not vary in time. The important small parameter k_0a is the focus of the perturbation expansion, where $k_0=\omega_0/c$ is the wavenumber of the incident plane-wave and where a is the rms value of the random surface height.

When a time-harmonic plane-wave with time-dependence $\exp(-i\omega_0 t)$ interacts with the time-varying rough surface depicted in Fig. 1, the spectrum of the scattered field is not restricted to the single excitation frequency ω_0. However, in the specialized case where the autocorrelation function of the statistically stationary rough surface can arise from a temporally static surface in uniform translation, simple coordinate transformation of well known static results from perturbation theory is applicable. The acoustic pressure field $\psi(x,y,t)$ is a function of the two spatial coordinates x and y, the time t, and is a solution to the source-free scalar wave equation. The Dirichlet boundary condition $\psi[x,a\zeta(x,t),t]=0$ applies to the total field, for all time, on the soft surface $y_s=a\zeta(x,t)$. The scattered field, which is defined here as the difference between the total field and the incident plane-wave that exists everywhere in the absence of the impenetrable surface, exhibits radiative behavior as $y\to+\infty$.

By expressing the soft surface as $y_s=a\zeta(x,t)$, where the rms surface height a appears explicitly, the mean-square value of the normalized random surface $\zeta(x,t)$ is unity. This real-valued two-dimensional stochastic process (or field) is assumed to be wide-sense stationary in both its spatial (x) and temporal (t) dependencies. It has zero mean-value $E\{\zeta(x,t)\}=0$, and the specific autocorrelation function adopted for the present study is

$$E\{\zeta(x,t)\zeta(x',t')\}=R(x-x',t-t')=\exp\left\{-\left[\frac{(x-x')-u(t-t')}{b}\right]^2\right\}. \quad (1)$$

This is a simpler version of a surface autocorrelation function proposed by Medwin and Clay.[6] Initial studies that incorporated their additional cosine factor were complex enough to warrant a reduction in the number of parameters. The stochastic scattered field caused by a rough surface described by the simple function (1) is rich enough in physical

a)Author to whom correspondence should be addressed. Electronic mail: rscharst@bama.ua.edu

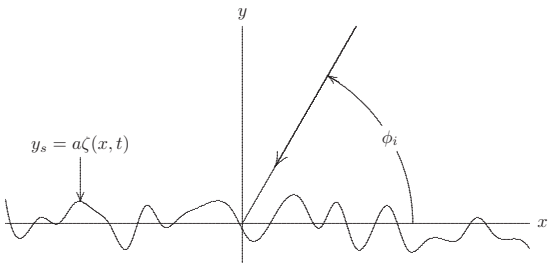


FIG. 1. Plane-wave incident upon a time-varying, rough surface.

features to capture our present attention while it maintains the dominant features of the time-varying rough surface. The parameter u (m/s) is a "group velocity" in that the Gaussian function in Eq. (1) propagates in the x direction with speed u. The parameter b (m) is a correlation length of the random surface.

A time-harmonic plane-wave of unit amplitude and frequency ω_0 is incident from the angle ϕ_i measured from grazing. In the absence of the roughness, that is, for $a=0$, the total field is the sum of the incident and reflected plane-waves

$$\psi_0(x,y,t) = e^{-ik_0(x\cos\phi_i+ct)}[e^{-ik_0y\sin\phi_i} - e^{ik_0y\sin\phi_i}] = -2i\sin(k_0y\sin\phi_i)e^{-ik_0(x\cos\phi_i+ct)}, \quad (2)$$

where $k_0=\omega_0/c$ is the exciting wavenumber with c the sound speed. This is the zeroth-order or geometrical acoustics field. The common perturbation expansion for the total field in powers of the rms surface height a, normalized to the original source wavelength in that $k_0a=2\pi a/\lambda_0$, is

$$\psi(x,y,t) = \psi_0(x,y,t) + k_0a\psi_1(x,y,t) + (k_0a)^2\psi_2(x,y,t) + \cdots. \quad (3)$$

Terms higher in order than $(k_0a)^2$ are an unnecessary complication because the case of small k_0a is sufficiently interesting and manageable. Perturbation theory and the philosophy of the second-order correlation approach to the stochastic problem are nicely treated by Rytov *et al.*[7] and by Sobczyk.[8] Interpretation and dynamic evolution of the coherence function are detailed by Beran and Parrent.[9] Warnick and Chew[10] gave a condensed summary of the voluminous literature on various approaches to rough surface scattering: An observation of the importance of analytic methods in building physical insight is still true. The paper by Shaw *et al.*[5] is similar in spirit to the work here, except that their use of finite source distances requires the evaluation of particular integrals. The uniform plane-wave is nice because finite source effects do not clutter the results.

II. FIRST-ORDER PERTURBATION FIELD

The first-order perturbative field is forced by the nonzero boundary data

$$\psi_1(x,0,t) = -\frac{\zeta(x,t)}{k_0}\frac{\partial}{\partial y}\psi_0(x,0,t) = 2i\sin(\phi_i)\zeta(x,t)e^{-ik_0(x\cos\phi_i+ct)}, \quad (4)$$

and it inherits its zero mean $E\{\psi_1(x,y,t)\}=0$ from that of $\zeta(x,t)$. The autocorrelation function of the complete $\psi_1(x,y,t)$ can be worked out from scratch using multiple Fourier integrals in all of the independent variables. However, because the autocorrelation function (1) of the rough surface can arise as a simple translation

$$E\{\zeta(x-ut)\zeta(x'-ut')\} = R[(x-x')-u(t-t')], \quad (5)$$

where

$$E\{\zeta(x)\zeta(x')\} = R(x-x') \quad (6)$$

defines the wide-sense statistical stationarity of a temporally static surface, a proper coordinate transformation as given by Morse and Ingard,[11] for example, can be applied directly to the standard results of perturbation theory for the static surface. These are summarized by Sobczyk,[8] for example. When applied to a Fourier integral representation for the first-order scattered field, the coordinate transformation exhibited in Eq. (5) gives

$$E\{\psi_1(x,y,t)\psi_1^*(x',y',t')\} = \frac{2b}{\sqrt{\pi}}\sin^2\phi_i\exp\left[-i\omega_0\left(1+\frac{u}{c}\cos\phi_i\right)(t-t')\right] \cdot \int_{-\infty}^{\infty} d\alpha\exp\left\{-\frac{1}{4}[(\alpha+k_0\cos\phi_i)b]^2\right\} \cdot \exp\{i[\alpha(x-x')+\beta y-\beta^* y'-u\alpha(t-t')]\}. \quad (7)$$

The separation constant β that accompanies the y-dependence, or the y-component of the wavenumber, is defined by the usual separation equation

$$\alpha^2+\beta^2 = k^2, \quad (8)$$

but now

$$k = k_0 + \frac{u}{c}(\alpha+k_0\cos\phi_i). \quad (9)$$

The coherence function of $\psi_1(x,y,t)$ is wide-sense stationary in x and t since only the differences $x-x'$ and $t-t'$ appear in Eq. (7). The evanescent contributions that manifest when β is imaginary are nonstationary in y. But for both y and y' sufficiently large, such exponentially decreasing contributions to the integral are negligible and the first-order perturbation field is effectively wide-sense stationary in all three of the independent variables x, y, and t.

The demarcation between evanescence and propagation, according to Eq. (8), occurs when $k^2=\alpha^2$. The two solutions to this quadratic are

$$\alpha_+ = \frac{k_0\left(1+\frac{u}{c}\cos\phi_i\right)}{1-\frac{u}{c}} \quad (10)$$

$$\alpha_- = -\frac{k_0\left(1+\frac{u}{c}\cos\phi_i\right)}{1+\frac{u}{c}}. \tag{11}$$

Hence, under the restriction $y=y'$ big enough to neglect the nonstationary evanescent contributions, the "far-field" coherence function of the first-order perturbation field is

$$E\{\psi_1(x,y,t)\psi_1^*(x',y',t')\} \approx \frac{2b}{\sqrt{\pi}}\sin^2\phi_i \exp\left[-i\omega_0\left(1+\frac{u}{c}\cos\phi_i\right)(t-t')\right] \cdot \int_{\alpha_-}^{\alpha_+} d\alpha \exp\left\{-\frac{1}{4}[(\alpha+k_0\cos\phi_i)b]^2\right\} \cdot \exp\{i\alpha[(x-x')-u(t-t')]\}. \tag{12}$$

Because of the wide-sense stationarity of the surface and the uniformity of the exciting plane-wave, the horizontal coordinate x and time t are grouped together in Eq. (12) the same as in the surface coherence function (1), as in

$$s = x - x' - u(t-t'). \tag{13}$$

Required then is the integral

$$I = \int_{\alpha_-}^{\alpha_+} d\alpha \exp\left\{-\frac{1}{4}[(\alpha+k_0\cos\phi_i)b]^2\right\} e^{i\alpha s}. \tag{14}$$

Introduction of the change-of-variable

$$z = \frac{1}{2}(\alpha+k_0\cos\phi_i)b - is/b \tag{15}$$

permits the integral (14) to be expressed in terms of the error function

$$I = \frac{\sqrt{\pi}}{b}\exp[-ik_0 s\cos\phi_i]\exp\left[-\left(\frac{s}{b}\right)^2\right][\mathrm{erf}(B)-\mathrm{erf}(A)], \tag{16}$$

with

$$A = -\frac{k_0 b\sin^2\frac{1}{2}\phi_i}{1+\frac{u}{c}} - i\frac{s}{b}, \tag{17}$$

$$B = \frac{k_0 b\cos^2\frac{1}{2}\phi_i}{1-\frac{u}{c}} - i\frac{s}{b}. \tag{18}$$

The error function of complex argument is computed using the algorithm of Weideman.[12] The resultant specialized ($y=y'\gg 1$) coherence function (12)

$$E\{\psi_1(x,y,t)\psi_1^*(x',y,t')\} = 2\sin^2\phi_i \exp\left[-i\omega_0\left(1+\frac{u}{c}\cos\phi_i\right)(t-t')\right] \cdot \exp[-ik_0 s\cos\phi_i]\exp\left[-\left(\frac{s}{b}\right)^2\right][\mathrm{erf}(B)-\mathrm{erf}(A)] \tag{19}$$

explicitly displays the dependence upon all of the geometrical and acoustical parameters of the physical problem.

For large values of the independent variable $s\to\infty$, a growth factor in the error functions tends to cancel the rapid decay in the Gaussian function. Calculations and insight are improved by a large s-expansion. It is more efficient to work directly with the original integral I rather than grappling with the asymptotics of the separate error functions. A single integration-by-parts applied to Eq. (14) yields a single-term asymptotic expansion

$$I(s) = \int_{\alpha_-}^{\alpha_+} d\alpha f(\alpha)e^{i\alpha s} = \left.\frac{f(\alpha)e^{i\alpha s}}{is}\right|_{\alpha=\alpha_-}^{\alpha_+} - \frac{1}{is}\int_{\alpha_-}^{\alpha_+} d\alpha f'(\alpha)e^{i\alpha s}. \tag{20}$$

Algebra that exploits the details exhibited by Eqs. (17) and (18) yields the large s form

$$isI(s) = \exp\left[-\frac{(k_0 b)^2\cos^4\frac{1}{2}\phi_i}{\left(1-\frac{u}{c}\right)^2}\right]\exp\left[i\frac{\left(1+\frac{u}{c}\cos\phi_i\right)}{1-\frac{u}{c}}k_0 s\right] - \exp\left[-\frac{(k_0 b)^2\sin^4\frac{1}{2}\phi_i}{\left(1+\frac{u}{c}\right)^2}\right]\cdot\exp\left[-i\frac{\left(1+\frac{u}{c}\cos\phi_i\right)}{1+\frac{u}{c}}k_0 s\right]. \tag{21}$$

Note that the magnitude of the associated coherence function ultimately decays as $|s|^{-1}$.

III. SECOND-ORDER PERTURBATION FIELD

The second-order field $\psi_2(x,y,t)$ in Eq. (3) is forced by the boundary data from the first-order field, as in

$$\psi_2(x,0,t) = -\frac{\zeta(x,t)}{k_0}\frac{\partial}{\partial y}\psi_1(x,0,t). \tag{22}$$

A single integral representation for the expectation of the second-order perturbation field is

$$E\{\psi_2(x,y,t)\} = \frac{b\sin\phi_i}{\sqrt{\pi}k_0}\exp[-i(k_0 x\cos\phi_i + k_0 ct)] \cdot \int_{-\infty}^{\infty} d\alpha\beta e^{i\beta y}\exp\left\{-\frac{1}{4}[(\alpha+k_0\cos\phi_i)b]^2\right\}, \tag{23}$$

where Eqs. (8) and (9) still apply. The x and t behavior is

identical to that of the incident plane-wave. A stationary phase approximation of the integral, as $y \to \infty$, yields

$$E\{\psi_2(x,y,t)\} = \frac{k_0 b}{\sqrt{2k_0 y}} \sin\phi_i \frac{\left(1+\frac{u}{c}\cos\phi_i\right)^{3/2}}{\left[1-\left(\frac{u}{c}\right)^2\right]^{5/4}} \cdot \exp\left\{-\frac{1}{4}\left[\frac{k_0 b\left(\frac{u}{c}+\cos\phi_i\right)}{1-\left(\frac{u}{c}\right)^2}\right]^2\right\} \cdot \exp[-i(k_0 x\cos\phi_i + k_0 ct + \pi/4)] \cdot \exp\left\{ik_0 y \frac{1+\frac{u}{c}\cos\phi_i}{\left[1-\left(\frac{u}{c}\right)^2\right]^{1/2}}\right\}. \quad (24)$$

IV. DISCUSSION OF RESULTS

To order $(k_0 a)^2$ the mean value of the perturbation expansion (3) is

$$E\{\psi(\mathbf{r},t)\} = \psi_0(\mathbf{r},t) + (k_0 a)^2 E\{\psi_2(\mathbf{r},t)\}, \quad (25)$$

where ψ_0 is the deterministic, geometrical acoustics field (2) and where the large y behavior of the mean value of ψ_2 is given by Eq. (24). Recall from the development that the mean value of ψ_1 is zero. As $u/c \to 1^-$ the stationary phase approximation of the integral in Eq. (23) requires adjustment, but for reasonable values of normalized speed, the information in the approximation (24) is substantial. This perturbation term in the distant mean field (25) is proportional to $y^{-1/2}$ and it inherits the x and t behavior of the incident wave. The Gaussian factor in Eq. (24) shows that $E\{\psi_2\}$ rapidly dissipates with the important parameter $k_0 b$. Large $k_0 b$ is the high frequency regime where the correlation length b of the rough surface interacts with a substantial number of the incident acoustic wavelengths, with the net effect that the superposition of the random contributions to the scattered field tends to cancel on the average. However, if $k_0 b$ is small enough so that the Gaussian factor is closer to unity, then the leading $k_0 b$ in Eq. (24) becomes the primary limiting factor. The strength of the specular component ψ_2 is a strong function of the incidence angle ϕ_i. In the more interesting case of large $k_0 b$, the amplitude factor in Eq. (24) becomes highly peaked around its maximum at $\cos\phi_i = -u/c$, where, in an average sense, the incident wavefront is essentially keeping up with the surface fluctuations.

The autocorrelation or coherence function of the total field is, again up to order $(k_0 a)^2$,

$$E([\psi(\mathbf{r},t) - E\{\psi(\mathbf{r},t)\}][\psi^*(\mathbf{r}',t) - E\{\psi^*(\mathbf{r}',t')\}]) = (k_0 a)^2 E\{\psi_1(\mathbf{r},t)\psi_1^*(\mathbf{r}',t')\}. \quad (26)$$

Hence, the coherence function (7) of the first-order perturbation field is the central descriptor of the scattering from the rough surface. With the independent space/time variable s defined in Eq. (13), introduce the abbreviated notation

$$|\rho(s/b)| = |E\{\psi_1(x,y,t)\psi_1^*(x',y,t')\}| \quad (27)$$

for the magnitude of the large $y = y'$ coherence function. The grouping of the field coordinates x, x', t, and t' occurs naturally in the physical mathematics, as in Eq. (13), and therefore we follow nature's lead and keep these coordinates together as they appear. It is entirely expected that the ratio s/b of the appropriately organized space-time coordinates to the rough surface correlation length is the important parameter in the scattering problem.

The variation in $|\rho|$ with normalized s/b is graphed in Figs. 2–4 for $u/c = 0$, 0.5, and 0.9, respectively, when the

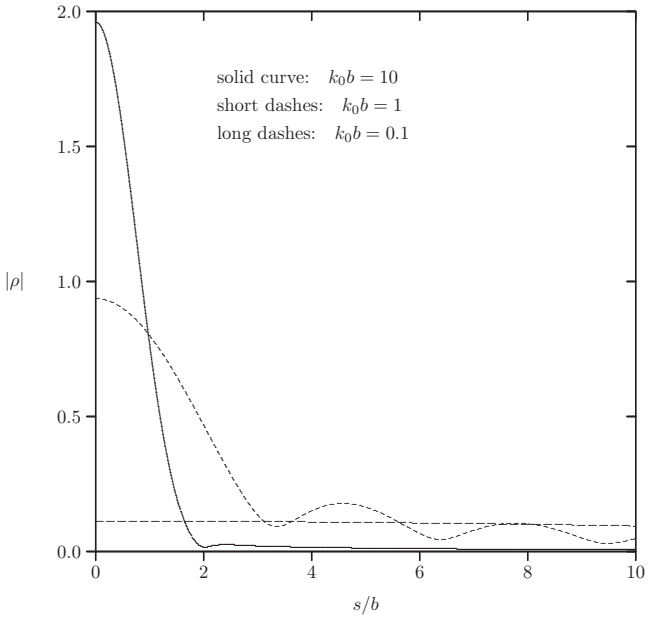


FIG. 2. Coherence function of first-order perturbation field. Case: $u/c = 0$ and $\phi_i = \pi/4$; $s = x - x' - u(t - t')$ and $y = y' \gg 1$.

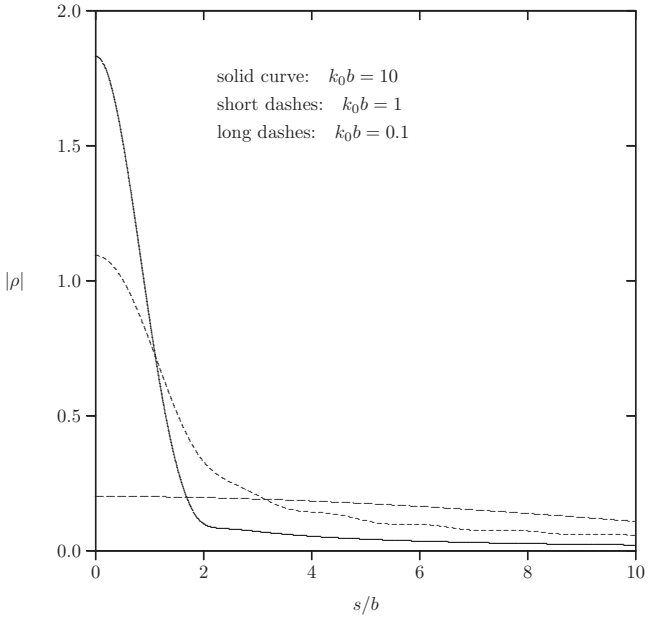


FIG. 3. Coherence function of first-order perturbation field. Case: $u/c = 0.5$ and $\phi_i = \pi/4$; $s = x - x' - u(t - t')$ and $y = y' \gg 1$.

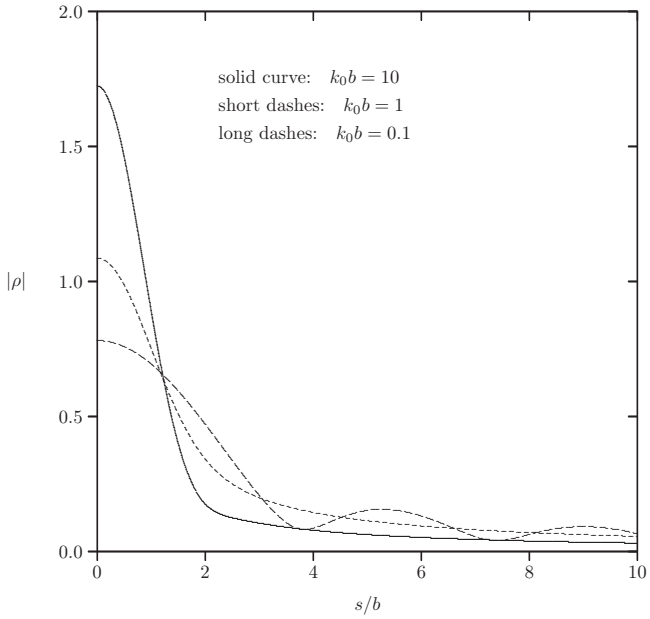


FIG. 4. Coherence function of first-order perturbation field. Case: $u/c = 0.9$ and $\phi_i = \pi/4$; $s = x - x' - u(t - t')$ and $y = y' \gg 1$.

exciting plane-wave is incident from the representative angle $\phi_i = \pi/4$. The three curves in each of Figs. 2–4 correspond to low, intermediate, and high frequencies as far as the correlation length b is concerned. The selected values $k_0b = 1/10$, 1, and 10 cover a wide range of frequencies.

Figure 2 shows that when the rough surface group velocity is zero, the first-order coherence function is a rather peaked function of s/b for the high frequency ($k_0b = 10$), compared to the essentially flat variation at the low frequency ($k_0b = 0.1$), that is substantially lower (about 5%) in magnitude. As u/c is increased, first to 0.5 in Fig. 3 and then to 0.9 in Fig. 4, the lower frequency graphs of $|\rho|$ display behavior more like that of the higher frequency $|\rho|$ from the static ($u/c = 0$) surface. A faster group velocity u in the rough surface correlation function (4) lets the incident acoustic wave experience more interactions with the roughness per wavelength. Therefore, if the acoustic correlation length k_0b is relatively small, the effect of increasing the group velocity u/c is similar to starting with a larger correlation length. It is "per wavelength" because both the exciting plane-wave and the scattering surface are infinite in extent, and the natural length scale that has physical meaning for the wave is its inherent spatial period.

This feature, that is so apparent in the graphs of Figs. 2–4, of increasing u/c mimicking a larger k_0b, is also obvious in the mathematics. The effect is clearly seen in the arguments A and B, defined in Eqs. (17) and (18), respectively, of the error functions in Eq. (19). The physical problem of Fig. 1 is symmetric about $\phi_i = \pi/2$, allowing for the sign of u/c. If u/c approaches +1, or simply increases positively, then the effect is the same as increasing k_0b in the parameter B of Eq. (18). Similarly, if u/c approaches −1, then the parameter A of Eq. (17) experiences an effectively larger k_0b.

The correlation function (19) and the graphical display of its behavior in Figs. 2–4 is the product of the analysis reported here and is, of course, specific to the assumed surface correlation function (1). Although rather restricted and idealized, such analytical results build insight into the stochastic scattering physics and are potentially useful in interpreting statistical data generated by Monte Carlo simulation using purely numerical models.

V. CONCLUSIONS

The coherence function for the random field scattered by the time-varying, slightly-rough surface is proportional to $(k_0a)^2$, where $k_0 = \omega/c$ is the wavenumber of the incident acoustic wave and a is the rms value of the rough surface. The scattered field is wide-sense stationary in both time and in the coordinate that is parallel to the rough surface. Far from the surface, the scattered field is also approximately wide-sense stationary in the normal coordinate. Two important parameters are k_0b, where b is the correlation length of the rough surface, and the normalized group velocity u/c. If the surface group velocity is zero and if k_0b is small, say, $k_0b \lesssim 1$, the random component of the scattered field is diffuse and rather uninteresting. Acoustically large correlation lengths, such as $k_0b \gtrsim 10$, produce a scattered field that is significantly correlated over a surface correlation length b, regardless of the surface group velocity. If k_0b remains small but the surface group velocity becomes a sizable fraction of the acoustic velocity, say, $u/c \gtrsim 1/2$, then the scattered field is noticeably imprinted with the dispersive characteristics of the rough surface and the correlation function maintains its cohesion and propagates similar to the case of larger k_0b.

ACKNOWLEDGMENTS

This work has in part been supported by the Office of Naval Research through the Naval Research Laboratory (Program Element No. 61153N-32). This document (NRL/JA/7180-09-0041) has been reviewed and approved for public release.

[1] B. E. Parkins, "Coherence of acoustic signals reradiated from the time-varying surface of the ocean," J. Acoust. Soc. Am. **45**, 119–123 (1969).
[2] C. S. Clay and H. Medwin, "Dependence of spatial and temporal correlation of forward-scattered underwater sound on the surface statistics. I. Theory," J. Acoust. Soc. Am. **47**, 1412–1418 (1970).
[3] E. Y. Harper and F. M. Labianca, "Perturbation theory for scattering of sound from a point source by a moving rough surface in the presence of refraction," J. Acoust. Soc. Am. **57**, 1044–1051 (1975).
[4] D. R. Dowling and D. R. Jackson, "Coherence of acoustic scattering from a dynamic rough surface," J. Acoust. Soc. Am. **93**, 3149–3157 (1993).
[5] W. T. Shaw, A. J. Dougan, and R. J. A. Tough, "Analytical expressions for correlation functions and Kirchhoff integrals for Gaussian surfaces with ocean-like spectra," IEEE Trans. Antennas Propag. **44**, 1454–1463 (1996).
[6] H. Medwin and C. S. Clay, *Fundamentals of Acoustical Oceanography* (Academic, Boston, 1998), pp. 589.
[7] S. M. Rytov, Yu. A. Kravtsov, and V. I. Tatarskii, *Principles of Statistical Radiophysics* (Springer-Verlag, Berlin, 1989), Vol. **4**, pp. 157–169.
[8] K. Sobczyk, *Stochastic Wave Propagation* (Elsevier, Amsterdam, 1985), pp. 201–207 and 214–217.
[9] M. J. Beran and G. B. Parrent, Jr., *Theory of Partial Coherence* (Prentice-Hall, Englewood Cliffs, NJ, 1964), pp. 36–44.
[10] K. F. Warnick and W. C. Chew, "Numerical simulation methods for rough surface scattering," Waves Random Media **11**, pp. R1–R30 (2001).
[11] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1986), pp. 717–727.
[12] J. A. C. Weideman, "Computation of the complex error function," SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **31**, 1497–1518 (1994).

Grating lobe reduction in transducer arrays through structural filtering of supercritical plates

Brian E. Anderson,[a] W. Jack Hughes, and Stephen A. Hambric
Applied Research Laboratory, The Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804

(Received 24 October 2008; revised 3 June 2009; accepted 4 June 2009)

The effect of placing a structural acoustic filter between water and the transducer elements of an array to help reduce undesirable grating lobes is investigated. A supercritical plate is mounted to transducer elements with a thin decoupling polyurethane layer between the transducers and the plate. The plate acts as a radiation/incidence angle filter to pass energy at angles near normal incidence, but suppress energy at large incidence angles. Grating lobe reduction is achieved at the expense of limiting the available steering of the main lobe. Within this steer angle limitation, the main lobe can be steered as normal while the grating lobe level is reduced by the plate's angular filtering. The insertion of a plate structural filter provides, in principal, an inexpensive and easily implemented approach to extend usable frequency bandwidth with reduced level grating lobes, without increasing the number of array elements. Even though the data match theory well, a practical material has yet to be found that possesses optimal material properties to make the proposed idea practical. This work represents the first attempt to advantageously utilize a plate above its critical frequency to provide angular dependent sound transmission filtering.
 [DOI: 10.1121/1.3159366]

PACS number(s): 43.38.Hz, 43.30.Yj, 43.40.Rj, 43.55.Rg [AJZ] Pages: 612–619

I. INTRODUCTION

According to traditional array theory, the usable frequency bandwidth of an array of regularly spaced point sources is limited according to its spacing.[1–3] The optimal spacing, d, is typically set equal to half the desired center frequency wavelength, λ, [Fig. 1(a) displays a point source line array with these geometrical properties identified]. Thus, the resulting array may be operated below and above the center frequency, within certain limitations. Below the center frequency, the radiated beam widens and approaches an omni-directional beam pattern at very low frequencies. Above the center frequency, the radiated beam narrows. At frequencies greater than or equal to twice that of the center frequency, grating lobes appear in the radiated beam pattern as duplicates of the main beam of sound energy (i.e., duplicates in radiation level, but not in the direction radiated). These undesirable grating lobes result from an aliasing effect due to the discrete nature of an array of elements, the regular spacing of elements, and the relationship between the wavelength and inter-element spacing.

The above discussion assumes that the main energy is sent in a direction normal to the plane of the array. An array may also be used to steer sound radiation in different directions by introducing appropriate electronic phases to the array elements. When an array is steered, the grating lobe(s) appear in the beam pattern at frequencies lower than twice the center frequency. Grating lobes in radiation beam patterns cause operational confusion for technology that relies on arrays, such as medical ultrasound and underwater SONAR. In transmit mode, grating lobes cause an array to transmit sound energy in unintended directions. In receive mode, grating lobes prevent an array's ability to detect the direction of incoming sound energy. However, larger frequency bandwidth allows higher resolution and greater ranging capabilities in applications where both transmit and receive modes are used.

There have been many different attempts to reduce or eliminate grating lobes in radiation patterns in SONAR array applications and in medical ultrasound array applications. SONAR arrays are often designed for high source level output in a narrow frequency band and are not traditionally used in the broad frequency range such as that used in the ultrasound community. Consequently, the majority of studies involving grating lobe reduction have traditionally come from the ultrasound community. The most commonly used techniques, aimed at grating lobe reduction in directivity patterns, attempt to break up the regular element spacing or regular element size which gives rise to grating lobes. Current techniques of reducing grating lobes include sparse element spacing,[4–17] annular rings of elements,[18–22] use of adjacent staggered arrays,[23–25] pseudorandom continuous wave (cw) signals,[26–28] and exploitation of harmonic imaging.[29]

There are many existing arrays with regularly spaced elements that have no practical means for reducing their grating lobes. None of the methods mentioned above are practical solutions to retrofit onto these array systems. The methods mentioned above either require changes to the original array design, require impractical modifications to the input signals, and/or are not practical solutions for underwater SONAR arrays. A technique is needed that directly eliminates grating lobes, which could be retrofitted onto existing array

[a]Author to whom correspondence should be addressed. Present address: Department of Physics and Astronomy, Brigham Young University, N283 ESC, Provo, Utah 84602. Electronic mail: bea@byu.edu

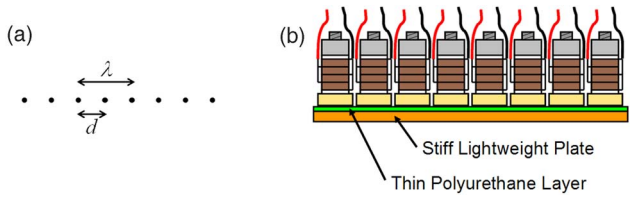


FIG. 1. (Color online) (a) Illustration of a point source line array. The variable d represents the center to center spacing between adjacent transducer elements. The variable λ represents the wavelength (in this case the center frequency wavelength used to design the array). (b) Illustration of a typical Tonpilz line array mounted onto a stiff lightweight plate. A thin layer of polyurethane was placed between the array and the plate.

designs. Grating lobe reduction would provide these existing arrays with better ranging capabilities and larger frequency bandwidth and therefore better image resolution compared to existing array capabilities. The goal of this paper is to present a new method to reduce grating lobes through a retrofitting approach.

The proposed technique for grating lobe level reduction presents a different approach to the problem than other methods. An extremely stiff and lightweight plate, operated at frequencies above its critical frequency, is placed in front of the transducer array to provide a radiation angle filter (and incidence angle filter) achieved through exploitation of the coincidence effect (a sound and structure interaction principle).[30] Figure 1(b) provides a simple illustration to show a typical eight element Tonpilz transducer line array that has been bonded onto a plate (or a bar in this case). The idea behind this technique was developed as an extension of the thesis work by Hutto,[31] Grosh,[32] and Phillips.[33]

A previous paper written by the authors showed that an array may steer sound energy through a supercritical plate without refraction and without widening the main lobe of energy, meaning that the plate does not alter the intended electronic phasing required to steer an array.[34] Sound energy passes through the plate at normal incidence and radiation angles less than the coincidence angle are attenuated somewhat due to the mass of the plate. However, sound energy at large radiation angles, greater than the coincidence angle, will be attenuated due to the stiffness of the plate. If the plate's material properties and geometry are selected appropriately and the array is operated within certain constraints, then grating lobe levels may be greatly reduced while main lobes are passed by the radiation angle filter. The insertion of a plate would provide an inexpensive and easily implemented approach to eliminating grating lobes without adjusting the number or positions of array elements.

This paper introduces a method to reduce grating lobes. However, the plate used for grating lobe reduction in this work does not provide a practical solution to the problem, as will be shown. While grating lobes are reduced by the plate used, additional relative increases in side lobe levels, combined with a reduction in the main lobe level, makes the plate used in this study an impractical solution to the problem. Thus this paper is intended to provide the blue print to design a passive structural acoustic filter to attenuate grating lobes when a plate material with sufficient stiffness to mass properties is developed.

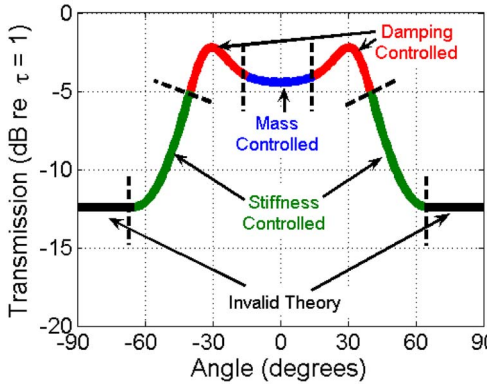


FIG. 2. (Color online) Example of the transmission through a supercritical plate which possesses a high stiffness to mass ratio at a single frequency versus angle of incidence.

II. THEORETICAL DEVELOPMENTS

A. Angular filtering

The critical or coincidence frequency of a plate is the frequency at which the bending wave speed in the plate equals the acoustic fluid wave speed. A high degree of sound transmission exists at grazing incidence for plates at their critical frequencies. Below this critical frequency, the sound transmission through the plate is governed by the well known mass law and is independent of incidence angle. However, above the critical frequency (the supercritical region), the transmission through the plate varies significantly with angle of incidence. Figure 2 shows an example of the transmission versus angle for a supercritical plate possessing a high stiffness to mass ratio. The mass of the plate governs the transmission at normal incidence and at angles near normal incidence. At an angle called the coincidence angle, θ_{CO}, the reactive part of the plate's impedance goes to zero, and the sound transmission is dominated by the amount of damping within the plate.[30] Thus, instead of a high degree of sound transmission at grazing incidence for the critical frequency, the high degree of sound transmission shifts to coincidence angles in the forward 180°. At angles greater than the coincidence angle, the transmission is dominated by the plate's stiffness.

A derivation of an expression for the angular dependent sound transmission coefficient, τ, through supercritical plates is given by Fahy and Gardonio[30] in their Eq. (5.38a):

$$\tau = \frac{(2\rho c \sec\theta)^2}{\left[2\rho c \sec\theta + \frac{1}{\omega} D\eta k^4 \sin^4\theta\right]^2 + \left[m\omega - \frac{1}{\omega} Dk^4 \sin^4\theta\right]^2}, \quad (1)$$

where ρ is the density of the acoustic fluid, c is the speed of sound in the acoustic fluid, θ is the angle of incidence, ω is the angular frequency, D is the plate's stiffness, η is the internal damping in the plate, k is the acoustic wavenumber in the fluid, and m is the mass per unit area in the plate. The transmission loss through the plate is

$$\mathrm{TL} = 20 \log_{10} \frac{1}{\tau}. \tag{2}$$

One may consider the coincidence angle to be the "cutoff angle" of the angular filter (similar to the cutoff frequency in frequency domain filtering). The range of angles between the two coincidence angles may be considered to be the angular pass band region. Finally, the angles between grazing incidence and the coincidence angles may be considered the angular stop band region. Fahy and Gardonio commented that Eq. (1) does not model the angular region ±78°–±90° (from normal incidence) near grazing incidence in practice[30] (at grazing incidence, the expression converges to a value of 1, regardless of the frequency used, meaning no loss in transmission). The authors came to a similar conclusion from their evaluations of experimental data; however, the authors have decided that the expression in Eq. (1) is invalid from grazing incidence to the first null in the expression that typically occurs at ±65° (see Fig. 2).

To use a plate as an angular sound filter, the plate's material properties and thickness should be chosen carefully to minimize attenuation of the pass band region and maximize the attenuation in the stop band region. Figure 2 is a good example of a filter which attenuates the stop band more than it attenuates the pass band. Another way to look at this angular filtering is that it increases the directivity in the forward direction, which may be beneficially exploited in other applications.

In selecting the optimum plate thickness, one must decide what coincidence angle, or cutoff angle, would be desired. In addition, one should decide the lowest frequency at which this cutoff angle is desired, keeping in mind that as frequency decreases the coincidence angle increases (thus the pass band region widens) according to the following equation:

$$\sin^2 \theta_{\mathrm{CO}} = \frac{c^2}{{c_B}^2} = \frac{c^2}{\omega} \sqrt{\frac{m}{D}}, \tag{3}$$

where c_B is the bending wave speed for a thin plate (thin plate theory ignores the effects due to shear deformation and rotary inertia). The parameters D and m may be written in terms of the plate's material properties and thickness as

$$D = \frac{Eh^3}{12(1-\sigma^2)}, \tag{4}$$

$$m = \rho_P h, \tag{5}$$

where E is the plate's Young's modulus of elasticity, h is the plate's thickness, σ is the plate's Poisson ratio, and ρ_P is the volumetric density of the plate. Equations (3)–(5) are adapted from Fahy and Gardonio's Eqs. (5.42)–(5.43). If one solves Eq. (3) for h and substitutes the desired cutoff angle with the selected frequency, then the interdependence of the thickness and the material properties may be examined (see Chap. 3 of the thesis by Anderson[35] for further details). Solving for h yields

$$h = \frac{c^2}{\omega \sin^2 \theta_{\mathrm{CO}}} \sqrt{\frac{12\rho_P(1-\sigma^2)}{E}}. \tag{6}$$

If a specific plate material is selected, then Eq. (6) may be used to calculate the thickness required to set the target coincidence angles at the desired operation frequency. As an example, the measured material properties of an alumina bar used in this study are $E=391$ GPa, $\rho_P=3956$ kg m^3, and $\sigma=0.22$. The desired coincidence angle was set to be $\theta_{\mathrm{CO}}=30°$ at a frequency of $f=50$ kHz. Thus the optimal plate thickness for alumina in this case is $h=9.7$ mm.

If Eq. (1) is evaluated at normal incidence, then one may estimate the expected pass band loss for an undamped ($\eta=0$) supercritical plate angular filter. The expected pass band transmission loss for an alumina plate of thickness $h=9.7$ mm is 12.3 dB. The stop band transmission loss may be determined at the minima of Eq. (1) that are located between grazing incidence and the coincidence angles. For the alumina plate, the minima are located at ±65° resulting in a maximum stop band transmission loss of 24.3 dB. Thus the supercritical plate may be used as a passive filter with a relative stop band transmission loss of 12 dB at large angles relative to the pass band region.

The directivity pattern, $H(\theta)$, without a plate may be calculated from typical line array theory (first term in curly brackets) multiplied by the single element directivity (second term in curly brackets),[1–3,35]

$$H(\theta) = \left\{ \frac{1}{N} \frac{\sin\left[\frac{Nkd}{2}(\sin\theta - \sin\theta_0)\right]}{\sin\left[\frac{kd}{2}(\sin\theta - \sin\theta_0)\right]} \right\} \times \left\{ \frac{\sin\left[\frac{1}{2}ka\sin\theta\right]}{\frac{1}{2}ka\sin\theta} \right\}, \tag{7}$$

where a is the width of a rectangular head transducer, and N is the number of array elements. The predicted angular dependent transmission coefficients [Eq. (1)] for an alumina plate were then multiplied by the without-plate directivity pattern to yield the with-plate directivity patterns. Figure 3 displays theoretical directivity patterns without and with the insertion of an alumina plate at 50 kHz. At an angle of −45° a reduction of 9 dB in the grating lobe level occurs due to the insertion of the plate. One may also notice the unfortunate relative increase in the transmission of side lobe levels centered at the predicted θ_{CO} (±30°) due to the high degree of sound transmission at these angles. The relative level increase at the coincidence angles depends on m since the depth of the pass band depends on m. This is one of the major issues in the use of a supercritical plate for angular filtering. The optimal plate material would have a low m while maintaining a high D/m. The thesis by Anderson[35] includes a discussion of ideal plate materials in Chaps. 3 and 7. It is important to note that the discussion given in Anderson's thesis pertains to underwater arrays with design fre-

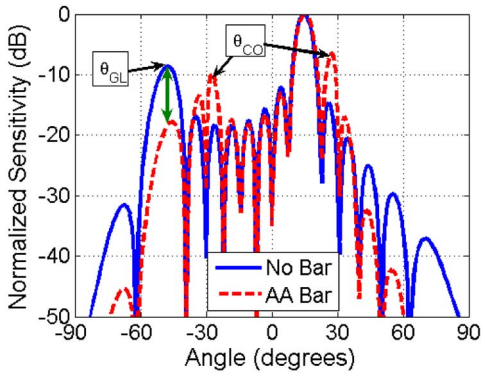


FIG. 3. (Color online) Example of grating lobe filtering through the addition of a supercritical plate. The plot displays theoretical sensitivity patterns at 50 kHz, steered to 15°, obtained from a point source directivity function with the addition of assumed single element directivity.

quencies of 25 kHz since the ideal material properties will depend on the desired frequency range.

B. Frequency response

An equivalent circuit will now be presented from which equations can be derived to model the frequency response of an array with and without the insertion of a plate. The mechanical impedance domain equivalent circuit for a plate loaded transducer in a receive condition is given in Fig. 4. A normal incidence plane wave with a first-order resistive loading approximation models the acoustic fluid medium. The incident wave impinges upon the plate, which is modeled by a waveguide T-network (the expressions used in T-networks are given by Mason on pp. 204–205 in Ref. 36). The thin layer of polyurethane between the plate and the transducer elements is also modeled by a waveguide T-network. Finally, a lumped-element equivalent circuit is given for the piezoelectric Tonpilz transducer.[2] The output potential signal of the equivalent circuit is proportional to the voltage signal measured at the transducer terminals.

A transfer function, $\mathrm{TF}_{\text{with plate}}$, between the input plane pressure wave, $\hat{p}$, and the output voltage signal, $\hat{V}$, may be derived to model the measured frequency response using standard circuit analysis techniques:

$$\mathrm{TF}_{\text{with plate}} = \frac{\hat{V}}{\hat{p}} = \frac{S}{\phi}\frac{Z_{P2}Z_{C2}Z_0}{Z_\alpha Z_\beta Z_\gamma}, \tag{8}$$

$$Z_\alpha = Z_A + Z_{P1} + Z_{P2}, \tag{9}$$

$$Z_\beta = Z_{P1} + Z_{P2} + Z_{C1} + Z_{C2}, \tag{10}$$

$$Z_\gamma = Z_{C1} + Z_{C2} + Z_{T1} + Z_{T0}, \tag{11}$$

where S is the surface area of a single transducer, ϕ is the electro-mechanical coupling factor, and Z represents a mechanical impedance quantity. The subscripts designate the following: resistive acoustic loading A, plate's masslike quantity $P1$, plate's compliancelike quantity $P2$, compliant layer's masslike quantity $C1$, compliant layer's compliantlike quantity $C2$, transducer's lumped mechanical model $T1$ (including mass, compliance, and resistance), and transducer's electrical capacitance $T0$. The transfer function for an unloaded transducer (no plate), $\mathrm{TF}_{\text{no plate}}$, is found by setting $Z_{P1}=Z_{P2}=Z_{C1}=Z_{C2}=0$ and $Z_{P2}=Z_{C2}=\infty$ and redoing the circuit analysis

$$\mathrm{TF}_{\text{no plate}} = \frac{\hat{V}}{\hat{p}} = \frac{S}{\phi}\frac{Z_A + Z_{T1} + Z_{T0}}{Z_{T0}}. \tag{12}$$

A complete discussion of this derivation and the expressions for the impedance terms are given in the thesis by Anderson.[35]

The transducers used in the measurements were characterized using electrical impedance measurements to extract average electro-mechanical lumped-element parameters. The geometry and material property values used for the alumina bar are $h=10.2$ mm (measured), $E=391$ GPa (measured), $\rho=3956$ kg/m^3 (measured), and $\sigma=0.22$ (manufacturer specification). These parameters, along with material properties of the plate and compliant layer, were used to evaluate the expressions for $\mathrm{TF}_{\text{with plate}}$ and $\mathrm{TF}_{\text{no plate}}$.

III. EXPERIMENT RESULTS

Experiments were conducted in a water tank that measures $5.3\times5.5\times7.9$ m^3 in size. The water tank is kept at a constant temperature and is lined with acoustical absorbing material. A calibrated transmitter or a calibrated receiver can be placed at one end of the tank to transmit/receive sound toward/from a transducer array under test (AUT) at the other end of the tank. A computer controlled turntable can rotate the AUT with fractional degree increments. When the AUT is in a receive mode, up to 64 channels of complex data may be simultaneously acquired by a vector signal analyzer. The advantage of receive mode measurements is that in the post-processing phase, the array can be arbitrarily steered to any angle by applying appropriate digital delays.

Figure 5 shows a photograph of the array module just before being lowered into the water tank. Two different array configurations were studied, one with a bar and one without. This paper uses the terms plate and bar almost interchangeably. The method is presented in the context of using a plate,

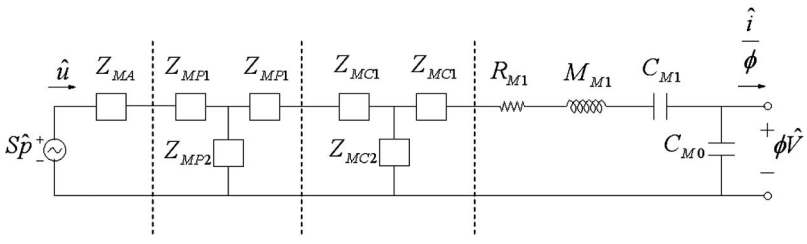


FIG. 4. Equivalent circuit model of an acoustically loaded and plate loaded Tonpilz piezoelectric transducer in a receive condition with waveguide circuits for the plate and compliant layers.

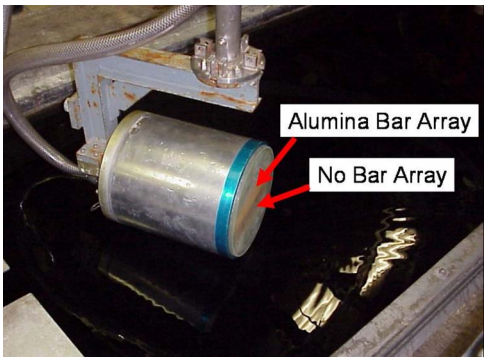


FIG. 5. (Color online) Photograph of the alumina bar array module used in experiments.

but the experiments utilized a bar. Both arrays consisted of transducer elements with nominal resonance frequencies of 25 kHz. Both arrays consisted of regularly spaced elements with a center to center spacing $d=30$ mm. According to standard line array theory, a grating lobe should be present for all steer angles, θ_0, at a frequency of 50 kHz. When the array is steered to $\theta_0=0°$, or broadside, grating lobes should appear at $\pm 90°$, according to point source line array theory. However, due to the fact that the elements have finite sizes, with square heads of length 29.5 mm, the directivity of a single element influences the overall directivity pattern according to the first product theorem (or product theorem).[3]

One array was a one-dimensional eight element Tonpilz transducer array that was mounted onto an alumina bar. A 1.6 mm thick layer of polyurethane separated the transducers from the bar. The alumina bar is a high purity 99.8% aluminum oxide ceramic (AmAlOx 68 alumina) and is 10.2 mm thick, 30.5 m wide, and 244 mm in length. Plastic shims were placed between neighboring transducers to maintain a consistent spacing between all transducers. The bar was then bonded onto a polyurethane acoustic window of an array module. A second eight-element line array was also bonded directly to the acoustic window without a bar between the array and the window to allow measurements on a similar array without the insertion of the bar. This array module is pictured in Fig. 5. Additional photos of this array module are found in Ref. 35.

A. Directivity pattern measurements

Directivity pattern measurements were measured with the AUT submersed in the water tank. A calibrated source transducer was placed at the other end of the tank and the AUT was oriented such that the sound energy incident upon the AUT was normal to the array. The calibrated source transducer was also used later in the frequency response measurements. The AUT was placed at a sufficient distance, r, away from the source transducer to assume that it was in the far field, $r=r_{\mathrm{FF}}$. The AUT was connected to a turntable that rotated the AUT slowly as the source projected sound energy. The individual element responses, $p_n(\theta)=A_n(\theta)e^{j\phi_n(\theta)}$, were collected at various angles, between $-90°$ and $+90°$ relative to normal incidence, as the AUT was rotated. In the postprocessing phase, each of the N elements' magnitude, $A_n(\theta)$, and phase, $\phi_n(\theta)$, could be summed with appropriate steering phases, $i_n\omega\tau_0$, where i_n is the column number and

$$\tau_0 = \frac{d \sin \theta_0}{c}, \tag{13}$$

where θ_0 is the desired steer angle. The combined AUT's directivity pattern for any given steer angle, $p_{\mathrm{AUT}}(\theta)$, is given by

$$p_{\mathrm{AUT}}(\theta)|_{r=r_{\mathrm{FF}}} = \sum_{n=1}^{N} A_n e^{j\phi_n} e^{j(i_n\omega\tau_0)}. \tag{14}$$

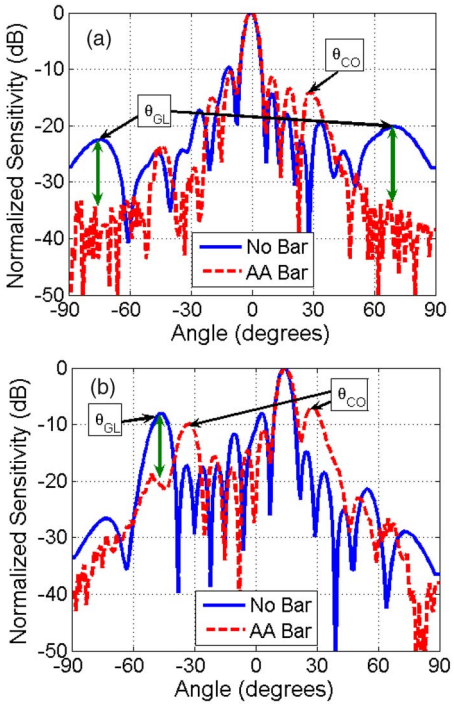


FIG. 6. (Color online) Measured far-field patterns for the alumina (AA) bar array with and without the bar. (a) 50 kHz and $\theta_0=0°$. (b) 50 kHz and $\theta_0=15°$.

Figure 6 displays the measured directivity patterns for both without-bar and with-bar conditions for the alumina bar array at 50 kHz. The patterns displayed in Fig. 6 are normalized with respect to their maxima to allow comparison of relative side lobe and grating lobe levels. The measured patterns for the alumina bar array show distinct peaks at the coincidence angles relative to the no-bar patterns, due to the high degree of sound transmission at the coincidence angles, as predicted in Sec. II. However, at 50 kHz for the alumina bar array, there is a clear reduction in level at the grating lobes. The alumina bar array patterns at 50 kHz reach the measurement system noise floor, approximately between -35 and -40 dB in the normalized plots given in Fig. 6, at large angles $>|\pm 60°|$; thus the true reductions in level at these large frequencies cannot be deduced.

Another convenient way to display the directivity pattern data is to plot the directivity patterns for various steer angles as slices in a three-dimensional surface plot.[34,35] This display format allows one to analyze the effect of the inser-

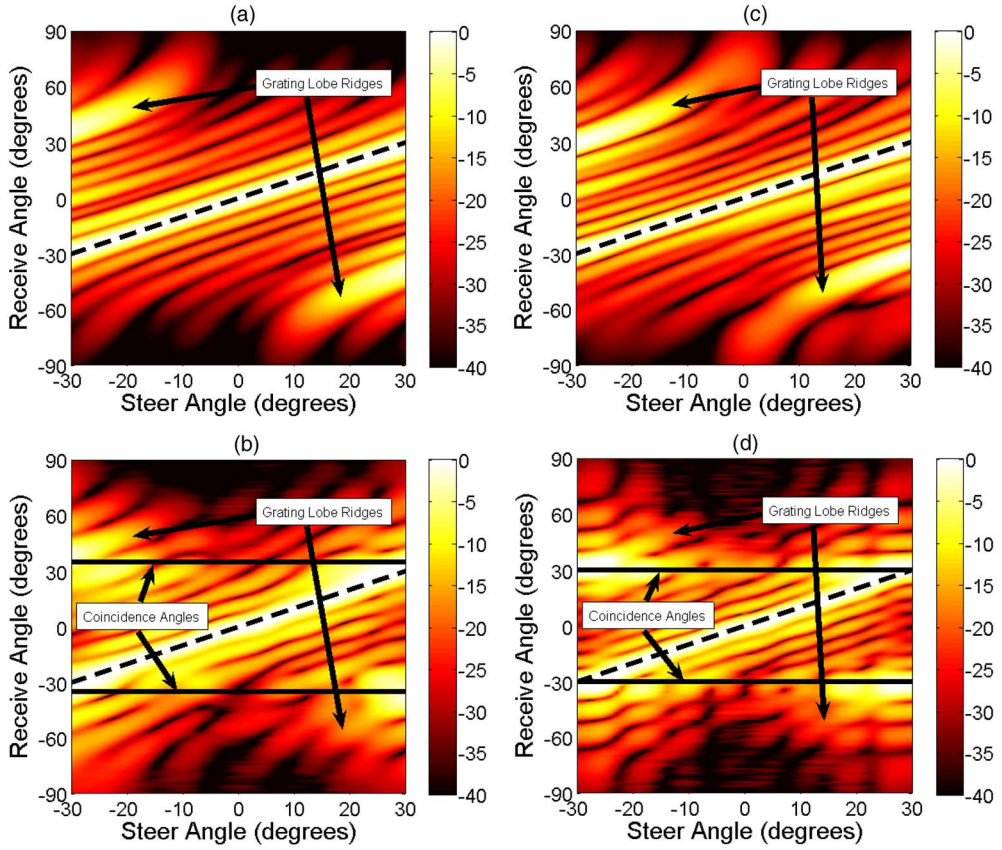


FIG. 7. (Color online) Directivity pattern surface plots for the alumina bar array at 45 kHz [(a) and (b)] and at 50 kHz [(c) and (d)] without the bar [(a) and (c)] and with the bar [(b) and (d)]. The main lobe ridge is identified by the black dashed lines. The grating lobe ridges are also identified. The coincidence angles are identified by the solid black lines.

tion of the bar/plate across many steer angles at the same time. Figure 7 displays directivity pattern surface plots for the alumina bar array, at frequencies of 45 and 50 kHz. The grating lobe ridge in both figures is reduced or at least broken up at steer angles less than ±20°. Also, there is an apparent increase in sensitivity at the expected coincidence angles indicated by the black horizontal lines in the figure.

B. Frequency response measurements and simulations

On-axis frequency response measurements were made with the AUT lowered into one end of the water tank. The total receive response from the AUT was recorded as the source transducer was swept up in frequency. Frequency response measurements were made for the AUT without a bar and then again with a bar and these results are given in Fig. 8. Equations (8) and (12) are evaluated and also plotted in Fig. 8. The presence of the alumina bar introduces a 5–10 dB transmission loss across the frequency range. Also, the insertion of the alumina bar reduces the 26 kHz resonance in the no-bar frequency response and tends to flatten out the overall response. The on-axis frequency response simulations predict trends found in the measured data, though they do not match the levels correctly. The reason for the model's overestimation of the measured data is unknown. Certain features of the model seem to be shifted down in frequency—note that the knees in the model's prediction at 18 and 34 kHz should match the knees in the measured data at 23 and 42 kHz. The equivalent circuit model includes some low frequency approximations, particularly in the model of the transducers, that explain the discrepancies above 50 kHz. While the equivalent circuit modeling shows some promise in modeling frequency response data, further work must be done to improve the modeling, such as using waveguide T-networks to model the transducer's components.

IV. CONCLUSION

This study represents the first attempt to advantageously utilize a plate to provide angular dependent sound transmis-

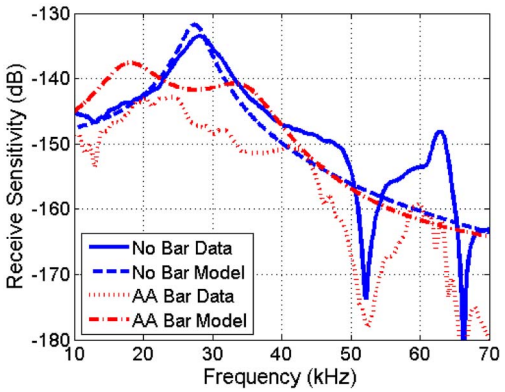


FIG. 8. (Color online) Frequency response comparison plots for the alumina (AA) bar array. The plot displays data and the equivalent circuit model results for the array with and without a bar in place.

sion filtering of array patterns above the plate's critical frequency (the supercritical frequency region). This study has shown that grating lobes can be reduced due to the insertion of a high stiffness to mass ratio supercritical plate. In addition to reducing grating lobes, the relative stop band transmission loss serves to increase the pattern directivity and suppress levels at large angles where flow noise can be problematic. Use of a plate as an angular filter does restrict steering between the positive and negative coincidence angles. The presence of the plate causes an insertion loss of the main lobe level and a relative increase in transmission in side lobes levels at the coincidence angles; both of these effects depend on the mass per unit area of the plate.

This study has found that while grating lobe levels can be significantly reduced, a practical material has yet to be found for the specific array configuration used. Typical reductions of 10 dB have been found in the grating lobe level (though the reduction depends on the steer angle of interest), while typical increases of 8 dB were found in the side lobe level at $\pm 30°$. The alumina material tested in this study did not provide optimal results from a practical standpoint, but the physics behind this technique has provided compelling evidence that this technique will succeed in reducing grating lobe levels when a plate is developed which possesses a higher stiffness to mass ratio than currently available materials.

ACKNOWLEDGMENTS

This research was sponsored by the Office of Naval Research Code No. 333, Dr. David Drumheller. In addition to ONR support for B.E.A., support was also provided by a Penn State University Graduate Research Fellowship, a Penn State College of Engineering Fellowship, an Audio Engineering Society Educational Foundation Award, and Penn State Applied Research Laboratory Exploratory and Foundational Research Program funding. The assistance of Paul Bednarchik, Courtney Burroughs, Bob Dashem, Mark Geleskie, Greg Granville, Bill Thomas, Dave Van Tol, and Mark Wilson of the Penn State Applied Research Laboratory is recognized and appreciated. Facilities located in the Applied Research Laboratory of The Pennsylvania State University were used in this work.

[1]V. M. Albers, *Underwater Acoustics Handbook II* (The Pennsylvania State University Press, State College, PA, 1965).
[2]D. Stansfield, *Underwater Electroacoustic Transducers* (Peninsula, Newport Beach, CA, 1991).
[3]D. T. Blackstock, *Fundamentals of Physical Acoustics* (Wiley, New York, 2000).
[4]O. T. von Ramm, S. W. Smith, and F. L. Thurstone, "Grey scale imaging with complex TGC and transducer arrays," Proc. Soc. Photo-Opt. Instrum. Eng. **70**, 266–270 (1975).
[5]S. W. Smith, H. G. Pavey, and O. T. von Ramm, "High speed ultrasound volumetric imaging system—Part 1: Transducer design and beam steering," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **38**, 100–108 (1991).
[6]G. R. Lockwood and F. S. Foster, "Optimizing sparse two-dimensional transducer arrays using an effective aperture approach," Proc.-IEEE Ultrason. Symp. **3**, 1497–1501 (1994).
[7]E. B. Hutchinson, M. T. Buchanan, and K. Hynynen, "Evaluation of an aperiodic phased array for prostate thermal therapies," Proc.-IEEE Ultrason. Symp. **2**, 1601–1604 (1995).
[8]W. J. Zehner, "SONAR array with reduced grating lobes," U.S. Patent 5,414,676 (1995).
[9]G. R. Lockwood, P. Li, M. O'Donnell, and F. S. Foster, "Optimizing the radiation pattern of sparse periodic linear arrays," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 7–14 (1996).
[10]E. B. Hutchinson and K. Hynyen, "Intracavitary ultrasound phased arrays for noninvasive prostate surgery," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 1032–1042 (1996).
[11]S. A. Goss, L. A. Frizzell, J. T. Kouzmanoff, J. M. Barich, and J. M. Yang, "Sparse random ultrasound phased array for focal surgery," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 1111–1121 (1996).
[12]L. R. Gavrilov, J. W. Hand, P. Abel, and C. A. Cain, "A method of reducing grating lobes associated with an ultrasound linear phased array intended for transrectal thermotherapy," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1010–1017 (1997).
[13]S. D. Sokka, R. L. King, N. J. McDonald, and K. H. Hynynen, "Design and evaluation of linear intracavitary ultrasound phased array for MRI-guided prostate ablative therapies," Proc.-IEEE Ultrason. Symp. **2**, 1435–1438 (1999).
[14]A. Trucco, "Aperture and element minimization in linear sparse arrays with desired beam patterns," Ultrasonics **38**, 161–165 (2000).
[15]G. T. Silva and M. Fatemi, "Linear array transducers for vibro-acoustography," Proc.-IEEE Ultrason. Symp. **1**, 629–632 (2002).
[16]Y. Liu and Z. Lin, "Active array beamforming using the frequency-response masking technique," Circuits and Systems, Proceedings of the 2004 International Symposium on Circuits and Systems, **3**, 197–200 (2004).
[17]O. Martinez, G. Godoy, M. A. G. Izquierdo, and L. G. Ullate, "Application of vernier thinning techniques to segmented annular arrays," Ultrasonics **42**, 977–982 (2004).
[18]S. Umemura and C. A. Cain, "Acoustical evaluation of a prototype sector-vortex phased-array applicator," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 32–38 (1992).
[19]O. Martinez, L. G. Ullate, M. A. G. Izquierdo, and T. Sanchez, "Incidence of element distribution on the ultrasonic field of segmented annular arrays," Ultrasonics **38**, 176–178 (2000).
[20]O. Martinez, L. G. Ullate, and A. Ibanez, "Comparison of CW patterns from segmented annular arrays and squared arrays," Sens. Actuators, A **85**, 33–37 (2000).
[21]K. A. Snook, T. A. Ritter, T. R. Shrout, and K. K. Shung, "Design of a high frequency annular array for medical ultrasound," Proc.-IEEE Ultrason. Symp. **2**, 1161–1164 (2001).
[22]O. Martinez, M. Akhnak, L. G. Ullate, and F. Montero de Espinosa, "A small 2D ultrasonic array for NDT applications," NDT & E Int. **36**, 57–63 (2003).
[23]B. P. Hildebrand and G. J. Posakony, "Method and apparatus for energizing an array of acoustic transducers to eliminate grating lobes," U. S. Patent 4,179,683 (1979).
[24]J. R. Talman and G. R. Lockwood, "Theoretical evaluation of a 50 MHz split aperture linear phased array," Proc.-IEEE Ultrason. Symp. **2**, 1675–1678 (1997).
[25]F. J. Pompei and S. C. Wooh, "Phased array element shapes for suppressing grating lobes," J. Acoust. Soc. Am. **111**, 2040–2048 (2002).
[26]F. Dupenloup, J. Y. Chapelon, D. Cathignol, and O. Sapozhnikov, "The use of broadband signals to reduce grating lobe effects in HIFU tissue ablation," Proc.-IEEE Ultrason. Symp. **3**, 1865–1868 (1994).
[27]J. Y. Chapelon, F. Dupenloup, H. Cohen, and P. Lenz, "Reduction of cavitation using pseudorandom signals," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 623–625 (1996).
[28]F. Dupenloup, J. Y. Chapelon, D. J. Cathignol, and O. A. Sapozhnikov, "Reduction of the grating lobes of annular arrays used in focused ultrasound surgery," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 991–998 (1996).
[29]A. Bouakaz, C. T. Lancee, P. J. A. Frinking, and N. de Jong, "Simulations and measurements of nonlinear pressure field generated by linear array transducers," Proc.-IEEE Ultrason. Symp. **2**, 1511–1514 (1999).
[30]F. Fahy and P. Gardonio, *Sound and Structural Vibration, Radiation, Transmission and Response*, 2nd ed. (Academic, Oxford, U.K., 2007).
[31]F. Hutto, "Wave number-frequency response of free-free Timoshenko beams with multiple sources," MS thesis, The Pennsylvania State University, State College, PA (1986).
[32]K. Grosh, "An experimental system for measuring the wave number-frequency response of Timoshenko beams," MS thesis, The Pennsylvania State University, State College, PA (1988).
[33]J. Phillips, "The wave number-frequency filtering characteristics of compliant layers," MS thesis, The Pennsylvania State University, State Col-

lege, PA (1989).
[34]B. E. Anderson, W. J. Hughes, and S. A. Hambric, "On the steering of sound energy through a supercritical plate by a nearfield transducer array," J. Acoust. Soc. Am. **123**, 2613–2619 (2008).
[35]B. E. Anderson, "Grating lobe reduction in transducer arrays through structural filtering of supercritical plates," Ph.D. thesis, The Pennsylvania State University, State College, PA (2006); electronic access available at http://etd.psu.edu (Last viewed June, 2009).
[36]W. P. Mason, *Electromechanical Transducers and Wave Filters* (Van Nostrand, New York, 1942).

Influence of the anisotropy on zero-group velocity Lamb modes

Claire Prada[a] and Dominique Clorennec
Laboratoire Ondes et Acoustique, ESPCI-Université Paris 7-CNRS, UMR 7587, 10 rue Vauquelin, 75231 Paris Cedex 05, France

Todd W. Murray
Department of Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

Daniel Royer
Laboratoire Ondes et Acoustique, ESPCI-Université Paris 7-CNRS, UMR 7587, 10 rue Vauquelin, 75231 Paris Cedex 05, France

(Received 19 March 2009; revised 9 June 2009; accepted 11 June 2009)

Guided waves in a free isotropic plate (symmetric S_n and antisymmetric A_n Lamb modes) exhibit a resonant behavior at frequencies where their group velocity vanishes while their phase velocity remains finite. Previous studies of this phenomenon were limited to isotropic materials. In this paper, the optical generation and detection of these zero-group velocity (ZGV) Lamb modes in an anisotropic plate is investigated. With a circular laser source, multiple local resonances were observed on a silicon wafer. Experiments performed with a line source demonstrated that the frequency and the amplitude of these resonances depend on the line orientation. A comparison between experimental and theoretical dispersion curves for waves propagating along the [100] and [110] directions of the silicon crystal verified that these resonances occur at the minimum frequency of the S_1 and A_2 Lamb modes. Simulations indicated that it is possible to deduce the three elastic constants of the plate material with good accuracy from these measurements.
 [DOI: 10.1121/1.3167277]

PACS number(s): 43.40.Dx, 43.35.Yb, 43.20.Mv [YHB] Pages: 620–625

I. INTRODUCTION

Lamb waves are frequently used in the ultrasonic characterization of thin plates. Most approaches proposed in the literature use the dispersion characteristics of these waves to extract the mechanical properties or thickness of a given plate. Laser-based ultrasonic techniques are particularly attractive for exciting and detecting Lamb waves due to the fact that remote measurements can be made without the need for a coupling media.[1,2] In the simplest case, a pulsed laser source focused to a point or a line is used to excite broad band Lamb waves that are detected after a given propagation distance using an optical interferometer.[3–6] Lamb waves of a given wavelength can be excited by the interference of two laser sources, creating a sinusoidal intensity pattern on the sample surface.[7,8] The laser source may also be scanned over the sample surface to preferentially launch Lamb waves with a phase velocity matching the velocity of the moving source.[9] In both cases, the phase velocity is measured as a function of frequency to extract material property information, and a moderate propagation distance is required to measure the velocities with sufficient accuracy.

Recently, a new approach that allows for local measurements of the elastic properties or thickness of various plates was demonstrated. It is based on the excitation by a laser source of a sharp resonance at the minimum frequency of the first symmetric S_1 Lamb mode, where the group velocity of the mode goes to zero while the phase velocity remains finite. The laser source can be either an amplitude-modulated laser that generates Lamb waves at a particular frequency[10] or a pulsed laser that excites the resonant response in the time domain.[11] The plate vibration can then be measured at the excitation point on the sample surface using an optical interferometer. The remarkably high quality factor at megahertz frequencies of these zero group velocity (ZGV) resonances can be used for mapping nanometer scale thickness variations in thin plates as well as for the local measurement of material properties.[12] In addition to laser-based excitation, Holland and Chimenti[13] observed a strong transmission of air-coupled ultrasound at the S_1 ZGV resonance frequency in millimeter thick plates. Gibson and Popovics[14] also demonstrated that resonance observed in concrete during impact-echo testing corresponds to the S_1 ZGV Lamb mode.

Previous experimental studies of the S_1 ZGV resonance phenomena have been limited to isotropic materials. In this paper, the authors extend this analysis to anisotropic materials and study the local, transient response of a thin silicon plate to pulsed laser heating. Anisotropic plates, often comprised of either single crystal or composite materials, are employed for a multitude of applications in, for example, the aerospace, defense, microelectronic, and medical device industries.[15] The ZGV resonance differs markedly in anisotropic media in that the ZGV resonance frequency itself shows directional dependence. The excitation laser source shape can be selected to excite a band of resonance frequen-

[a]Author to whom correspondence should be addressed. Electronic mail: claire.prada-julia@espci.fr.

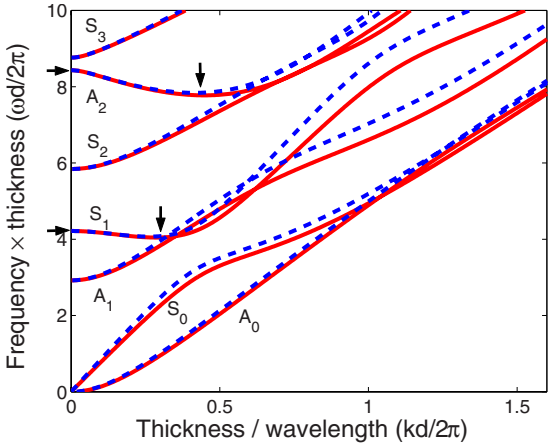


FIG. 1. (Color online) Dispersion curves in the (ω,k)-plane for Lamb waves propagating in a silicon plate of thickness d along the [100] axis (solid lines) and the [110] axis (dashed lines). Vertical arrows correspond to ZGV modes, and horizontal ones to thickness resonance frequencies.

cies over all propagation angles in the sample or to excite a narrow ZGV resonance in a given direction. A laser line source, for example, can be used to generate a sharp ZGV resonance in the direction perpendicular to the line. In this case, the ZGV resonance frequency can be "tuned" by rotating the excitation laser line source with respect to the sample. Laser generated ZGV resonance can be utilized for the local and noncontact nondestructive characterization of anisotropic plates.

II. ZGV LAMB MODES

The propagation of Lamb waves (frequency f, wavelength λ) is represented by a set of dispersion curves in the angular frequency $(\omega=2\pi f)$ and the wave number $(k=2\pi/\lambda)$ plane.[16] It is judicious to use variables normalized to the plate thickness d, such as the frequency thickness product $fd=\omega d/2\pi$ and the thickness to wavelength ratio $d/\lambda=kd/2\pi$. A silicon crystal of cubic symmetry is characterized by three elastic constants: $c_{11}=165.6$ GPa, $c_{12}=63.9$ GPa, $c_{44}=79.5$ GPa, and a mass density $\rho=2329$ kg/m^3. Figure 1 shows, for a [001]-cut silicon plate, the dispersion curves of the lower order symmetric and antisymmetric modes calculated with the software DISPERSE.[17] Curves were calculated for propagation directions along the [100] axis (azimuth angle $\varphi=0°$) and the [110] axis ($\varphi=45°$). For other directions, dispersion curves of a given mode lie inside the bundle delimited by these curves. The lowest order symmetric (S_0) and antisymmetric (A_0) modes originate at zero frequency, while all higher modes originate at a cut-off frequency f_c associated with either the longitudinal (V_L) or shear (V_T) wave velocity in the [001]-plate thickness direction:

$$V_L=\sqrt{c_{11}/\rho}=8433 \text{ m/s},$$

$$V_T=\sqrt{c_{44}/\rho}=5843 \text{ m/s}. \quad (1)$$

It can be observed that for high order modes, the group velocity $V_g=d\omega/dk$ vanishes at $k=0$, giving rise to a thickness resonance at the cut-off frequency. In addition, the group velocity of the first symmetric (S_1) and second antisymmetric (A_2) Lamb modes also vanishes for finite values of the wave number and wavelength $(k_0=2\pi/\lambda_0)$. At these ZGV points the acoustic energy, which cannot propagate in the plate, is trapped in the vicinity of the source. For the S_1 ZGV mode: $\lambda_0=3.50d$ and $3.80d$ in the [100] and [110]-directions, respectively. Thus, the vibration at the minimum frequency f_0 is localized in a region of dimension $\lambda_0/2$ approximately equal to twice the plate thickness. While this behavior is similar to that of an isotropic material,[12] an important distinction is that f_0 is dependent on the propagation direction. It has been shown on isotropic plates that the resulting local resonance can be efficiently excited by a laser pulse and optically detected by a sensitive interferometer.[10]

III. EXPERIMENTAL RESULTS

Experiments were carried out on a commercially available silicon wafer of thickness $d=525\pm5$ μm and diameter equal to 125 mm. Lamb waves were generated in the thermoelastic regime by a Q-switched Nd:YAG (yttrium aluminum garnet) laser (optical wavelength $\Lambda=1064$ nm) providing pulses having a 20-ns duration and 4-mJ of energy. The spot diameter of the unfocused beam is equal to 1 mm. Lamb waves were detected by a heterodyne interferometer[18] equipped with a 100-mW frequency doubled Nd:YAG laser (wavelength $\Lambda=532$ nm). This interferometer is sensitive to any phase shift along the path of the optical probe beam reflected by the moving surface. The contribution due to the local vibration of the plate is proportional to the component u of the mechanical displacement in the x-direction normal to the surface:

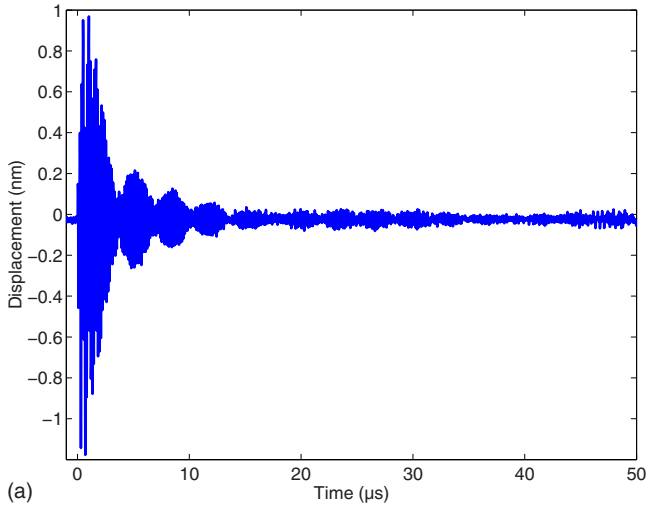
$$\Delta\phi_u=\frac{4\pi}{\Lambda_0}u. \quad (2)$$

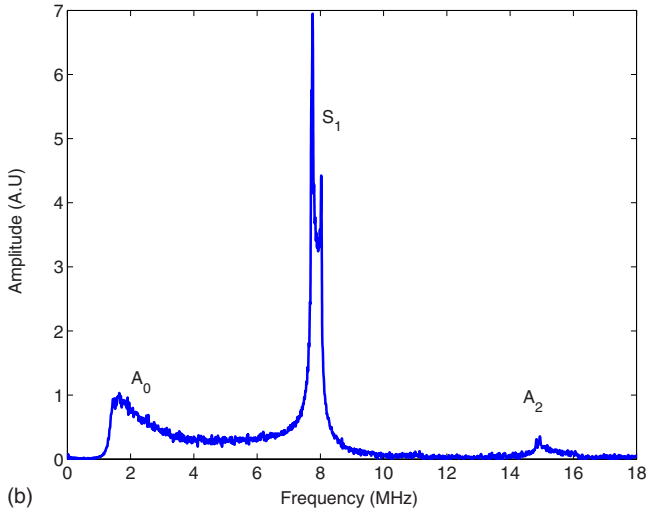
In many experiments the source and detection points are superimposed. The laser energy absorption heats the air in the vicinity of the surface and produces a variation Δn of the index along the optical path of the probe beam $(0<x<L)$. The resulting phase shift:

$$\Delta\phi_n=\frac{4\pi}{\Lambda_0}\int_0^L \Delta n(x)dx \quad (3)$$

induces a very large low frequency voltage, which saturates the electronic detection circuit. This spurious thermal effect is eliminated by interposing a high-pass filter having a cut-off frequency equal to 1 MHz. The calibration factor for mechanical displacement normal to the surface (10 nm/V) was constant over the detection bandwidth (1–20 MHz). Signals detected by the optical probe were fed into a digital sampling oscilloscope and transferred to a computer.

Figure 2(a) shows that the first 50 μs of the acoustic signal exhibit a series of maxima and minima regularly spaced by 3.3 μs, which reveals the existence of frequency beats. The frequency spectrum in Fig. 2(b), obtained from the whole signal (300 μs), is composed of two parts: a spread spectrum in the low-frequency range, corresponding to the non-resonant A_0 mode, and several sharp peaks. Taking into account the minimum frequency of the S_1 and A_2 branches in Fig. 1 and the nominal plate thickness (d





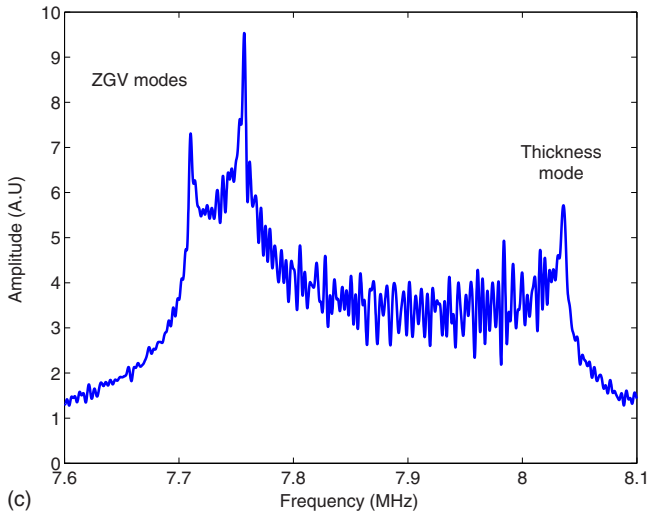


FIG. 2. (Color online) Point laser source. (a) Signal generated by the pulsed laser on a thin silicon plate and detected at the same point by the optical probe. (b) Frequency spectrum of the optically measured displacement from 0 to 18 MHz. (c) Expanded view of the S_1 ZGV mode resonances and thickness mode resonance.

$=0.525$ mm), peaks around 8 and 15 MHz can be ascribed to the ZGV resonance of S_1 and A_2 modes, respectively. Since the spectrum of the incident laser pulse is limited to 18 MHz, the excitation of the A_2 mode is less efficient than that of the S_1 mode. An expanded view of the frequency spectrum in the vicinity of the S_1 mode is shown in Fig. 2(c). The central part is composed of three sharp peaks. The quality factor $Q=2500$, deduced from the relative bandwidth ($Q=f/\Delta f$), is limited by the acquisition time. The peak at 8.03 MHz corresponds to the thickness resonance at the S_1 mode cut-off frequency $f_c=V_L/2d$, with $V_L=8433$ m/s and $d=0.525$ mm. In previous experiments on metallic plates,[11] the thickness mode resonance was not observed. Since silicon is partially transparent to near infrared radiation, the absorption of the YAG laser energy below the plate surface favors the generation of longitudinal waves and thus the excitation of the thickness extensional vibration. The amplitude modulation of the acoustic signal results from the beating of the S_1 thickness and ZGV modes. From the theoretical values of the minimum frequencies,

$$f_{0[100]}d = 4.044 \text{ MHz mm},$$

$$f_{0[110]}d = 4.069 \text{ MHz mm}, \quad (4)$$

for the [100] and [110] propagation directions, respectively, it appears that the other two peaks at 7.71 and 7.75 MHz correspond to the ZGV resonances for the S_1 mode in these directions. The laser source thus seems to preferentially excite local resonance for Lamb modes propagating in these directions. In order to confirm this hypothesis, the authors conducted another experiment where the point laser source used for Lamb wave excitation was replaced with a narrow line source.

A beam expander ($\times 7.5$) and a cylindrical lens (focal length 200 mm) were used to enlarge the laser beam and focus it to a narrow line on the plate surface. The optical energy distribution was close to a Gaussian and the absorbed power density was below the ablation threshold. The full length of the source at $1/e$ of the maximum value was found to be 12 mm and the width was estimated to be 0.3 mm. The plate was rotated from $\varphi=0°-360°$ in 1° steps and for each position the mechanical displacement was measured in the middle of the line. Figure 3(a) is a B-scan of the spectrum from 7.7 to 8.1 MHz. As expected, the thickness resonance frequency is independent of the line orientation. The ZGV resonance frequency, on the other hand, oscillates with a 90° period from the minimum value at $\varphi=0°$ to the maximum value at $\varphi=45°$. A 50-kHz frequency tuning can be achieved by rotating the line source.

The polar plot in Fig. 3(b) shows the amplitude of the ZGV resonance as a function of angle (φ) and maxima are observed at $\varphi=0°$ and 45° (and crystallographically equivalent directions). The authors assume that this effect could be explained by the deviation of the acoustic energy vector. In an anisotropic medium, the acoustic ray is normal to the slowness surface and, generally, is not collinear to the wave vector k. Figure 4 shows the slowness curve, i.e., the variations of the inverse of the S_1 ZGV phase velocity:

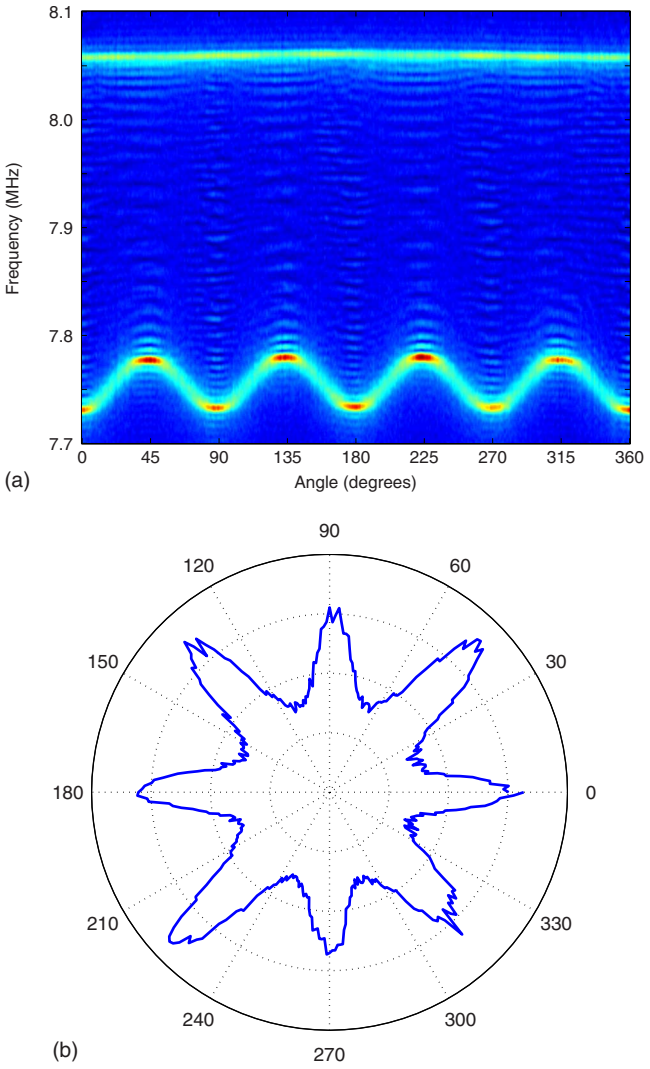


FIG. 3. (Color online) Line laser source. (a) *B*-scan of the spectrum measured as a function of the line orientation. The thickness resonance frequency is constant while the ZGV resonance frequency oscillates with a 90° period. (b) Polar plot (linear scale) of the resonance amplitude versus the propagation direction.

$$\frac{1}{V_0} = \frac{k_0}{2\pi f_0}, \tag{5}$$

plotted as a function of the propagation direction. The deviation of the energy velocity vector $\boldsymbol{V_e}$ from the wave vector

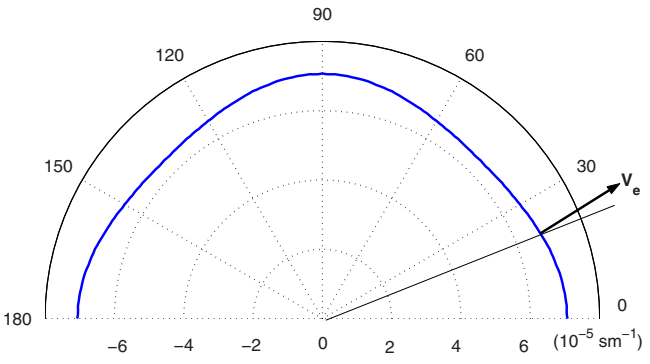


FIG. 4. (Color online) Slowness curve of the S_1 ZGV mode and deviation of the energy velocity vector $\boldsymbol{V_e}$.

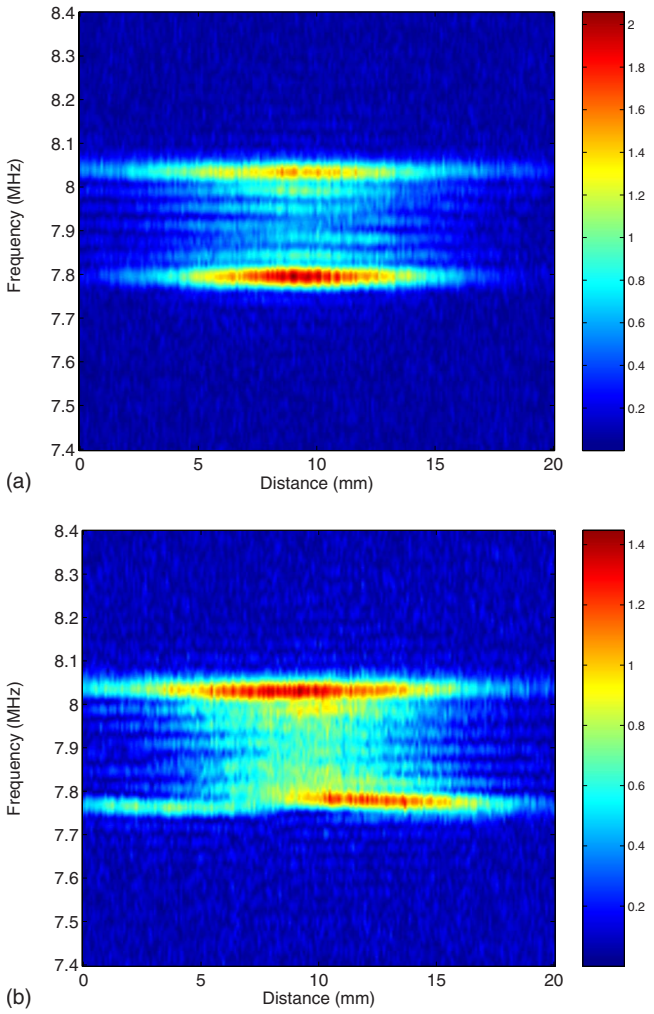


FIG. 5. (Color online) *B*-scan of the resonance amplitude measured along a line parallel to the source at a distance equal to 10 mm. (a) Line perpendicular to the direction $\varphi=0°$. (b) $\varphi=22.5°$: the energy distribution of the ZGV resonance is shifted from that of the thickness resonance mode.

vanishes for $\varphi=0°$ and 45° and reaches 10.1° for the direction $\varphi=22.5°$. In order to check that the acoustic beam is tilted, the amplitude of the resonances was measured along a line parallel to the source at a distance equal to 10 mm. Figure 5 is a frequency *B*-scan (±10 mm apart from the center of the line source) in the [100] and in the intermediate ($\varphi=22.5°$) directions. As the thickness resonance is not sensitive to the plate anisotropy, the upper line in the plot is not deviated from the normal to the line source. Conversely, the lower line in Fig. 5(b) shows that the energy distribution of the ZGV resonance is shifted by 2.2 mm, corresponding to a beam deviation angle equal to about 12.4°, slightly higher than the expected value.

It was previously demonstrated that the ZGV resonance results from the interference of the two waves generated in opposite directions with comparable amplitudes by the line source.[19] Owing to the negative slope of the S_1 mode dispersion curve for $k<k_0$ (Fig. 1), these two counter-propagating modes have the same (small) group velocity. The standing wave pattern is oriented along the propagation direction of the acoustic energy. In an anisotropic plate and in noncrys-

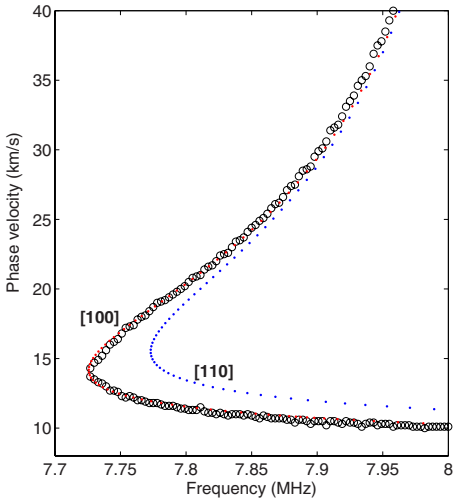


FIG. 6. (Color online) Phase velocity dispersion curves for a line source perpendicular to the [100] direction. Experimental points (circles) are compared with the theoretical curves computed for the S_1 Lamb mode propagating along [100]- and [110]-directions.

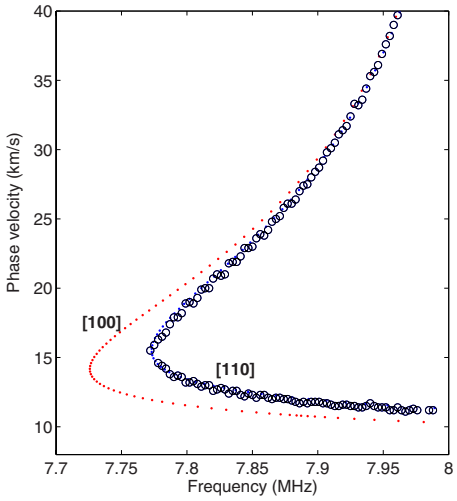


FIG. 7. (Color online) Phase velocity dispersion curves for a line source perpendicular to the [110] direction. Experimental points (circles) are compared with the theoretical curves computed for the S_1 Lamb mode propagating along [100]- and [110]-directions.

tallographic directions, the authors expected that the beam deviation reduces the efficiency of the interference and thus the amplitude of the ZGV resonance.

In order to identify the ZGV modes, they have measured the dispersion curves. For a given orientation of the line laser source, the distance r of the optical probe beam was varied from 0 to 12 mm in 0.1 mm steps. At each source to receiver distance, the normal displacement $u(r,t)$ was recorded during 100 μs with a 100 MHz sampling frequency. The measured signals are time Fourier transformed into $U(r,f)$. At each frequency, a spatial Fourier transform provides the spectrum in the time and spatial frequency domain:

$$\bar{U}(k,f) = \int U(r,f)e^{ikr}dr. \tag{6}$$

The spatial frequencies of the acoustic modes were determined by identifying the peaks in the power spectrum. Applying this procedure from 7.7 to 8 MHz allows us to plot the dispersion curves for the S_1 Lamb mode. The phase velocity dispersion curves for [100] and [110] propagation directions are shown in Figs. 6 and 7, respectively. Taken into account the small frequency range (300 kHz), experimental points are very close to the theoretical curves calculated for a plate thickness d=0.5235 mm. The maximum difference between theory and experiments (10 kHz) is on the order of the frequency resolution. As expected, the dispersion curves measured for other directions (φ=22.5°, for example) lie in between the curves calculated for [100]- and [110]-axis (Fig. 8).

The same signal processing was applied to the experimental data for frequencies around 15 MHz. Figure 9 shows that the high frequency peak in Fig. 2(b) can be ascribed to the minimum frequency of the A_2 Lamb mode. Due to the small signal to noise ratio, only the Fourier components close to the minimum frequency can be extracted from the experimental data. Nevertheless this ZGV mode is of great interest for the determination of the plate material properties. Simulations giving the shift in the resonance frequency for a 1% relative change of the elastic constants indicate that the S_1 ZGV Lamb mode is not sensitive to the variations of c_{12}. Conversely, the A_2 ZGV mode is very sensitive to the variations of c_{44} and relatively sensitive to a c_{12} change. The elastic constant c_{11} can be deduced directly from the thickness resonance frequency. Assuming a 15 kHz resolution in the frequency measurement, it is possible to determine c_{44} and c_{12} with 0.1% and 0.5% uncertainties, respectively, from the phase velocity dispersion curves in the minimum frequency region. The inversion of the dispersion data for a complete determination of the silicon elastic moduli is beyond the scope of this paper and may be accomplished following the approaches proposed in the literature.[20,21]

IV. CONCLUSION

The authors analyzed the mechanical response of a silicon plate to an excitation by a laser pulse focussed into a circular or a line source. The displacement normal to the

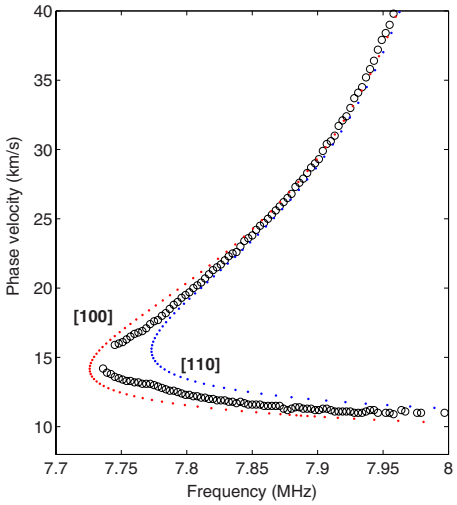


FIG. 8. (Color online) For a line source perpendicular to the 22.5° direction, experimental data (circles) lie in between the curves computed for the S_1 Lamb mode propagating along [100]- and [110]-directions.

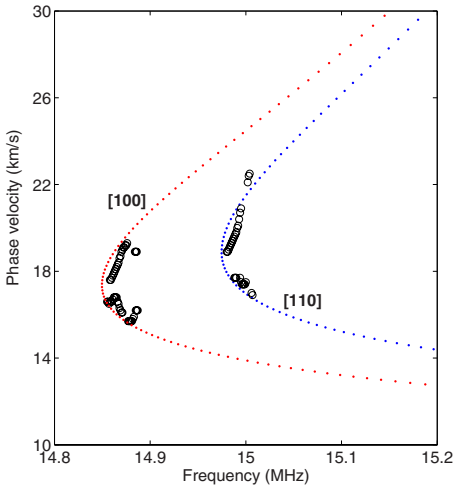


FIG. 9. (Color online) Phase velocity dispersion curves for a line source perpendicular to [100]- and [110]-directions. The experimental points (circles) are compared with the theoretical curves computed for the A_2 Lamb mode propagating along [100]- and [110]-directions.

plate surface was measured close to the source by a sensitive interferometer. This noncontact setup allows us to recover sharp peaks corresponding to the resonance of both thickness mode and ZGV Lamb modes. In contrast to the thickness resonance, ZGV resonances at the minimum frequency of Lamb mode dispersion curves are sensitive to the anisotropy of the plate material. Their amplitude and frequency depend on the orientation of the laser line source. In nonsymmetry directions and for a finite line source, the standing wave pattern resulting from the interference of the two counterpropagating waves is tilted from the normal to the line source. For the S_1 and A_2 ZGV Lamb modes, the minimum frequency is clearly identified from the comparison between experimental and theoretical dispersion curves. With an appropriate inversion algorithm, it should be possible to determine from these measurements the three elastic constants of the plate material with excellent accuracy.

[1]C. B. Scruby and L. E. Drain, *Laser Ultrasonics, Techniques and Applications* (Adam Hilger, New York, 1990).
[2]S. J. Davies, C. Edwards, G. S. Taylor, and S. B. Palmer, "Laser-generated ultrasound: Its properties, mechanisms and multifarious applications," J. Phys. D **26**, 329–348 (1993).
[3]L. F. Bresse, D. A. Hutchins, and K. Lundgren, "Elastic constant determination using generation by pulsed lasers," J. Acoust. Soc. Am. **84**, 1751–1757 (1988).
[4]S. E. Bobbin, J. W. Wagner, and R. C. Cammarata, "Determination of the flexural of thin films from measurements of the first arrival of the symmetric Lamb wave," Appl. Phys. Lett. **59**, 1544–1546 (1991).
[5]W. Gao, C. Glorieux, and J. Thoen, "Laser ultrasonic study of Lamb waves: Determination of the thickness and velocities of a thin plate," Int. J. Eng. Sci. **41**, 219–228 (2003).
[6]D. A. Hutchins, K. Lundgren, and S. B. Palmer, "A laser study of transient Lamb waves in thin materials," J. Acoust. Soc. Am. **85**, 1441–1448 (1989).
[7]J. A. Rogers and K. A. Nelson, "Study of Lamb acoustic waveguide modes in unsupported polyimide thin films using real-time impulsive stimulated thermal scattering," J. Appl. Phys. **75**, 1534–1556 (1994).
[8]J. A. Rogers, A. A. Maznev, M. J. Banet, and K. A. Nelson, "Optical generation and characterization of acoustic waves in thin films: Fundamentals and applications," Annu. Rev. Mater. Sci. **30**, 117–157 (2000).
[9]K. Yamanaka, Y. Nagata, and T. Koda, "Selective excitation of single-mode acoustic waves by phase velocity scanning of a laser beam," Appl. Phys. Lett. **58**, 1591–1593 (1991).
[10]C. Prada, O. Balogun, and T. W. Murray, "Laser-based ultrasonic generation and detection of zero-group velocity Lamb waves in thin plates," Appl. Phys. Lett. **87**, 194109 (2005).
[11]D. Clorennec, C. Prada, D. Royer, and T. W. Murray, "Laser impulse generation and interferometer detection of zero-group velocity Lamb modes," Appl. Phys. Lett. **89**, 024101 (2006).
[12]D. Clorennec, C. Prada, and D. Royer, "Local and noncontact measurement of bulk acoustic wave velocities in thin isotropic plates and shells using zero-group velocity Lamb modes," J. Appl. Phys. **101**, 034908 (2007).
[13]S. D. Holland and D. E. Chimenti, "High contrast air-coupled acoustic imaging with zero group velocity Lamb modes," Appl. Phys. Lett. **83**, 2704–2706 (2003).
[14]A. Gibson and J. S. Popovics, "Lamb wave basis for impact-echo method analysis," J. Eng. Mech. **131**, 438–443 (2005).
[15]B. Audoin, C. Bescond, and M. Deschamps, "Measurement of stiffness coefficients of anisotropic materials from pointlike generation and detection of acoustic waves," J. Appl. Phys. **80**, 3760–3771 (1996).
[16]D. Royer and E. Dieulesaint, *Elastic Waves in Solids* (Springer, Berlin, 1999), Vol. **1**.
[17]B. Pavlakovic and M. Lowe, DISPERSE software, v.2.0.16. Mechanical Engineering, Imperial College, London, 2005.
[18]D. Royer and E. Dieulesaint, in *Proceedings of the 1986 IEEE Ultrasonics Symposium* (IEEE, New York, 1986), p. 527.
[19]C. Prada, D. Clorennec, and D. Royer, "Local vibration of an elastic plate and zero-group velocity Lamb modes," J. Acoust. Soc. Am. **124**, 203–212 (2008).
[20]M. R. Karim, A. K. Mal, and Y. Bar-Cohen, "Inversion of leaky Lamb wave data by simplex algorithm," J. Acoust. Soc. Am. **88**, 482–491 (1990).
[21]S. I. Rokhlin and D. E. Chimenti, "Reconstruction of elastic constants from ultrasonic reflectivity data in a fluid coupled composite plate," Rev. Prog. Quant. Nondestr. Eval. **9B**, 1411–1418 (1990).

Long-term road traffic noise exposure is associated with an increase in morning tiredness

Yvonne de Kluizenaar[a)]
Department of Environment and Health, TNO, P.O. Box 49, 2600 AA Delft, The Netherlands and Department of Public Health, Erasmus University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

Sabine A. Janssen
Department of Environment and Health, TNO, P.O. Box 49, 2600 AA Delft, The Netherlands

Frank J. van Lenthe
Department of Public Health, Erasmus University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

Henk M. E. Miedema
Department of Environment and Health, TNO, P.O. Box 49, 2600 AA Delft, The Netherlands

Johan P. Mackenbach
Department of Public Health, Erasmus University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

(Received 11 September 2008; revised 13 May 2009; accepted 26 May 2009)

This study investigates the association between night time road traffic noise exposure (L_{night}) and self-reported sleep problems. Logistic regression was performed in a large population based cohort study (GLOBE), including over 18 000 subjects, to study the association between exposure at the dwelling façade and sleep problems. Measures of sleep problems were collected by questionnaire with two questions: "Do you in general get up tired and not well rested in the morning?" and "Do you often use sleep medication or tranquilizers?" After adjustment for potential confounders, a significant association was found between noise exposure and the risk of getting up tired and not rested in the morning. Although prevalence of medication use was higher at higher noise levels compared to the reference category ($L_{night} < 35$ dB), after adjustment for covariates this association was not significant. Long-term road traffic noise exposure is associated with increased risk of getting up tired and not rested in the morning in the general population. This result extends the earlier established relationship between long-term noise exposure and self-reported sleep disturbance assessed with questions that explicitly referred to noise and indicates that road traffic noise exposure during the night may have day-after effects.
 [DOI: 10.1121/1.3158834]

PACS number(s): 43.50.Qp, 43.50.Lj, 43.50.Rq [BSF] Pages: 626–633

I. INTRODUCTION

Undisturbed sleep is important for human functioning. During sleep, people recover mentally and physically from their activities and process information they have acquired during the day (Siegel, 2005; Marshall and Born, 2007; Drosopoulos *et al.*, 2007). Insufficient sleep is associated with feelings of sleepiness and fatigue during the daytime (Scott *et al.*, 2007; Elmenhorst *et al.*, 2008) and with decreased cognitive performance (Carter, 1996; Ouis, 1999; Raidy and Scharff, 2005, Murphy *et al.*, 2006). Furthermore, sleep deprivation has been associated with changes in physiological parameters, such as metabolic and endocrine function (Spiegel *et al.*, 1999), altered cardiac autonomic nervous system activity (Holmes *et al.*, 2002), and reduced acute immune system responses during stress (Wright *et al.*, 2007). One of the key features of subjective sleep quality is morning tiredness (Harvey *et al.*, 2008). Complaints of nonrestorative sleep (not feeling rested after sleep for at least three to four times a week) have been shown to be correlated with daytime impairment such as irritability and mental fatigue (Ohayon, 2005).

Environmental noise may disturb recuperation by activating the organism during sleep. Transportation is a prolific source of environmental noise during the night time in urban areas and has been identified as a major cause of sleep disturbance (Berglund *et al.*, 1999; Muzet, 2007). Exposure to transportation noise has been shown to induce both objectively measured and self-reported sleep disturbance (e.g., Pearsons *et al.*, 1995; Aasvang *et al.*, 2008; Basner *et al.*, 2006; Michaud *et al.*, 2007; Miedema and Vos, 2007). Considering the continuing growth of vehicular traffic and the large number of people exposed, disturbance of sleep by road traffic noise has become an increasingly important cause of concern.

[a)]Author to whom correspondence should be addressed. Electronic mail: yvonne.dekluizenaar@tno.nl

Effects of night time road traffic noise exposure on aspects of sleep have been found in both laboratory studies and in field studies with subjects exposed to habitual noise in their home situation (Passchier-Vermeer and Passchier, 2000; Franssen and Kwekkeboom, 2003; HCN, 2004). The observed effects include awakenings or sleep stage changes (Carter *et al.*, 1994), autonomic responses (Di Nisi *et al.*, 1990; Hofman *et al.*, 1995; Griefahn *et al.*, 2008; Graham *et al.*, 2009), body movements (Horne *et al.*, 1994; Passchier-Vermeer *et al.*, 2007), and self-reported noise-induced awakenings, difficulty falling asleep, and reduced sleep quality (Öhrström, 2000; Öhrström *et al.*, 2006a; Passchier-Vermeer *et al.*, 2007; Marks and Griefahn, 2007). While awakening reactions may be subject to habituation (Thiessen and Lapointe, 1983; Öhrström, 2000), traffic noise has been found to induce cardiac responses and motility in people who have lived in the vicinity of a major road for years (Hofman *et al.*, 1995; Passchier-Vermeer *et al.*, 2007). Furthermore, long-term effects of road traffic noise have been found on self-reported noise-related sleep disturbance and general sleep quality, as reported in questionnaires (HCN, 2004; Bluhm *et al.*, 2004), although few studies adjusted for potential confounders (Franssen and Kwekkeboom, 2003). On the basis of the pooled original data sets from 24 community surveys, exposure-effect relationships have been presented for the association between long-term night time transportation noise exposure and self-reported noise-related sleep disturbance (Miedema and Vos, 2007). In addition, some field studies have found next-day effects, such as tiredness in the morning and depressed mood as indicated in sleep logs, as well as poorer performance on reaction time tasks (Carter, 1996; Ouis, 1999; HCN, 2004).

Thus, there seems to be sufficient evidence that long-term traffic noise exposure is associated with self-reported noise-related sleep disturbance. However, little is known about the impact of long-term road traffic noise exposure on problems related to sleep such as morning tiredness and medication use. Although several field studies have been carried out (e.g., Öhrström, 1989; Öhrström and Skanberg, 2004; Passchier-Vermeer *et al.*, 2007), few epidemiological studies have assessed the relationship between long-term exposure to residential road traffic noise and sleep problems in the general population (e.g., Langdon and Buller, 1977; Kageyama *et al.*, 1997; Bluhm *et al.*, 2004). Furthermore, in previous studies, results may, to some extent, have been distorted by limitations in the study sample (e.g., non-random selection or relatively small sample), exposure assessment, or control for potential confounders. Moreover, in field studies and surveys designed specifically to investigate the community effects of noise, participants were usually aware of the noise focus of the study, and their response to questions concerning aspects of sleep may have been biased by their attitude toward the local road traffic exposure. The objective of the present study is to investigate the relationship between night time road traffic noise exposure (L_{night}) and self-reported sleep problems in a population based cohort study. As far as the authors know, this study is the first to investigate the relationship between night time road traffic noise exposure and morning tiredness and sleep medication use in such a large population based sample. In this study, objective measures are used for noise exposure, and odds ratios (ORs) are studied with adjustment for a broad spectrum of potential confounders. Furthermore, since the population study and questionnaire were not directed toward studying the effect of noise and noise exposure was determined independently, the subjects' attitude toward the local road traffic exposure is unlikely to have affected the results.

II. METHODS

A. Study population

The GLOBE study is a prospective cohort study carried out in the Netherlands, with the primary aim of explaining socio-economic inequalities in health. GLOBE is the Dutch acronym for Health and Living Conditions of the Population of Eindhoven and surroundings. Baseline data were collected in 1991. Details of the study protocol have been described elsewhere (Mackenbach *et al.*, 1994) and will only be briefly summarized here.

In 1991, an a-select sample (stratified by age, degree of urbanization, and socio-economic position) of 27 070 non-institutionalized subjects (aged 15–74 years) was drawn from 18 municipal population registers in the south-eastern part of The Netherlands and was asked to participate in the study. With a response rate of 70.1%, baseline information was collected from 18 973 individuals using a postal questionnaire. The area of study included the city of Eindhoven, which was the fifth largest city of The Netherlands in 1991.

B. Health outcome and covariates

The data collection comprised a broad range of potential confounders including sociodemographic variables (age, gender, marital status, and education), lifestyle factors [smoking, alcohol use, physical activity, and body mass index (BMI)], and living conditions (employment status and financial problems). Data for measures of sleep problems were available from the following questions in the questionnaire: "Do you in general get up tired and not well rested in the morning?" and "Do you often use sleep medication or tranquilizers?" The response format is "yes" or "no."

C. Noise exposure

The road traffic noise exposure of the subjects was calculated at the most exposed façade of the baseline home address with standard method SKM2 in accordance with requirements of the EU Environmental Noise Directive (END). For the analyses, the authors used the EU standard noise metric L_{night}. L_{night} (night level) is defined as the A-weighted "average" sound level (International Standards Organization, 2002) over a year during the period 23–7 h assessed at the façade of a dwelling with the highest overall exposure (i.e., most exposed façade). SKM2 is the sophisticated version of The Netherlands' standard method for noise modeling and producing noise maps in compliance with the END (VROM, 2006). SKM2 is implemented in Urbis (Borst and Miedema, 2005) that was used here for the exposure calculations. Noise calculations are carried out in two steps, calculating first the

emission and then the transmission. The emission calculations take into account traffic characteristics, including traffic intensities, traffic composition (percentage motorbikes, light duty, medium duty, and heavy duty vehicles), speed, road height, and road surface type. The transmission calculations take into account the distance between source (road) and dwelling façade, air attenuation, effects of (yearly) meteo conditions, ground attenuation, object screening, reflection of objects opposite the dwelling, and statistical diffraction for transmission. Noise exposure is calculated at the height of the center of the dwelling façade of the exposed subject. Very low noise exposure levels [below 35 dB(A)] were recoded as 35 dB(A) since this can be considered to be a lower limit of the night time ambient noise in most surroundings involved.

Input data for the noise emission calculations were a detailed digital map describing the geographic location of roads and the traffic characteristics for each road segment (including traffic intensities for each vehicle category, speed, and road surface type), provided by the local authorities of Eindhoven for the current situation (2004). Although traffic intensities may have increased, the road network is assumed to be rather stable, with only small (if any) but equal changes in noise exposure across the population. Traffic data were attached as attributes to the road segments for a dense network of roads, including highways, arterial roads, main streets, and principal residential streets.

Basis for the noise transmission calculations was digital maps with precise information on geographic location of buildings and ground characteristics [Topographic Service data (TOP10)] provided by The Netherlands Ministry of Housing, Spatial Planning and the Environment (VROM)/ Directoraat-Generaal Ruimte (DGR). Building height was derived from the Actual Height Information Netherlands, a 5×5 m^2 grid with height information based on laser altimetry. The geographic location of noise screens with their height was provided by the local authorities of Eindhoven. The geographical location of dwellings within the building contours [Topographic Service data (TOP10)] was identified with the use of address coordinates.

D. Statistical analysis

Logistic regression was performed to investigate the association between night time residential road traffic noise exposure (L_{night}) and self-reported sleep problems (getting up tired and not well rested in the morning and the use of sleep or tranquilizing medication). Estimated ORs are presented as approximation of relative risks, together with their corresponding 95% confidence intervals (CIs).

In the model, factors were included that were hypothesized *a priori* to potentially confound the relationship between traffic exposure and sleep problems. These variables are age, sex, BMI, physical activity, marital status, employment status, financial problems, alcohol use, smoking, and self-reported level of education. A P value of at most 0.05 was considered to be significant; a P value of 0.05–0.1 was considered an indication of a relationship.

Age was entered as a continuous variable, while gender, BMI, physical activity, marital status, employment status, financial problems, alcohol use, smoking, and education were entered as categorical variables. BMI (bodyweight divided by height squared) was categorized into four groups [underweight (BMI $<$ 20), normal weight range (BMI 20–25), overweight (BMI 25–30), and obese (BMI $>$ 30)]. Physical activity was available in four categories (none, little, moderate, and much physical activity). Marital status was categorized into four groups (married or living together, unmarried, divorced, widow/widower). Employment status was categorized in three categories, including "unemployed." Three categories of financial problems were distinguished (no difficulty, some difficulty, and large difficulty). Alcohol use was categorized into three groups (moderate, abstainer, and excessive). Data on smoking were available from the following question in the questionnaire: "Do you smoke?" The response format is as follows: "Yes," "No, but I have smoked in the past," "No, I never smoked," coded in three categories (current smoker, former smoker, and never smoker). Highest attained level of education was distinguished into four different categories (primary education, lower professional and intermediate general education, intermediate professional and higher general education, and higher professional education and university).

A sensitivity analysis was carried out to explore the effect of the inclusion of additional variables. In this analysis, a measure of occurrence of major life events, number of children living at home, and measures of cold or draught (answer categories: yes or no) and dampness (answer categories: yes or no) inside the dwelling were taken into account in the model in addition to age, sex, BMI, exercise, marital status, work situation, financial difficulties, smoking, alcohol use, and education. For major life events, a sum-score was used as the number of times respondents answered "yes" to one of nine questions on occurrence of major life events experienced during the last 12 months. These events included (1) moving house; (2) substantial decrease in financial situation; (3) being the victim of a serious crime (robbery, theft, physical abuse, or rape); (4) becoming unemployed; (5) partner or other family member (member of household) becoming unemployed; (6) serious disease of partner or family member (member of household) or parents

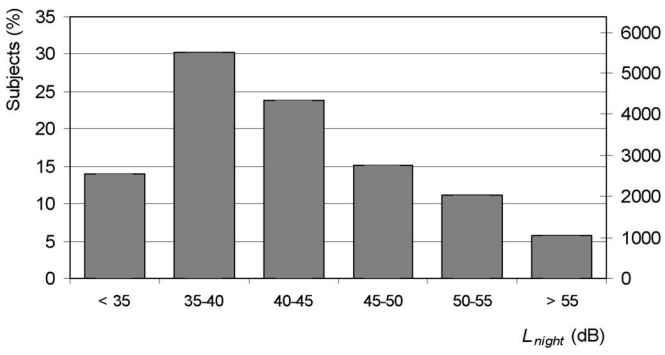


FIG. 1. Distribution of long-term average road traffic noise (L_{night}) (dB) exposure at the 1991 home address.

TABLE I. Characteristics of the GLOBE cohort by road traffic noise (L_{night}) (dB) exposure category (unadjusted).

L_{night} (dB)	<35	35–40	40–45	45–50	>50
N	2547	5514	4325	2742	3085
Age (year)	45.7	46.8	47.9	48.3	49.0
Sex: male (%)	49.6	48.3	48.3	47.6	48.3
BMI (QI>30) (%)	5.3	6.3	6.4	6.2	5.8
Physical activity: much (%)	35.1	33.0	32.7	33.1	30.8
Marital (married/living together) (%)	74.2	74.6	75.4	72.8	67.8
Work situation: unemployed (%)	9.9	10.4	10.8	8.7	10.9
Financial: much difficulty (%)	4.7	4.7	4.3	3.6	3.6
Smoking (%)	35.8	37.2	35.7	35.0	36.9
Alcohol use: excessive (%)	8.8	8.0	7.8	7.8	9.0
Education low (%)	18.3	21.0	22.3	20.5	22.6
Not rested in the morning (%)	17.9	19.4	20.3	21.0	20.2
Sleep/tranquilizing medication (%)	5.4	6.6	6.9	7.1	7.3

Variables are described by means and percentages in case of dichotomous variables.
Abbreviations are BMI (body mass index) and L_{night} (road traffic noise–night time noise level) (dB).

(in law); (7) death of partner; (8) death of parent (in law), child, brother of sister or close friend; and (9) divorce.

Missings in potential confounding variables (the percentage of missings for all confounding variables was below 5.6 %) were imputed, replacing the missing values with the most common category. All analyses were performed with SPSS (Version 11.0.1).

III. RESULTS

Figure 1 shows the distribution of subjects over night time road traffic noise exposure classes (L_{night}) for the GLOBE study sample at their 1991 home address. The spatial variation in road traffic noise exposure is substantial and shows a difference in exposure between the lowest and highest 5% of dwellings exceeding 20 dB (L_{night}), ranging from about 35 dB (urban background) to more than 55 dB (L_{night}) in the vicinity of roads.

Table I shows the distribution of the total study sample over the road traffic noise exposure categories. Before adjustment for potential confounders, the prevalence of both markers of sleep problems (getting up tired and not rested in the morning; use of sleep or tranquilizing medication) seems higher at higher night time noise levels (Fig. 2).

As shown in Table II and Fig. 3, in the GLOBE study sample an association was found between road traffic noise exposure (L_{night}) and getting up tired and not rested in the morning. Compared to the reference category (L_{night} <35 dB), the OR was higher in all higher noise exposure categories. The ORs were found to increase with increasing noise level but showed a slight decrease in the highest exposure category. Overall, an OR of 1.08 was found (95% CI: 1.02–1.14) per 10 dB increase in L_{night}. A sensitivity analysis was carried out to explore the effect of the inclusion of additional variables. Additional adjustment for major life events, and number of children living at home, and cold or draught and dampness inside the dwelling did not substantially affect the association for night time road traffic noise and getting up tired and not rested in the morning.

While the unadjusted results indicate that there may be an association between road traffic noise exposure and the use of sleep medication or tranquilizers, after adjustment for potential confounders this relationship was not significant.

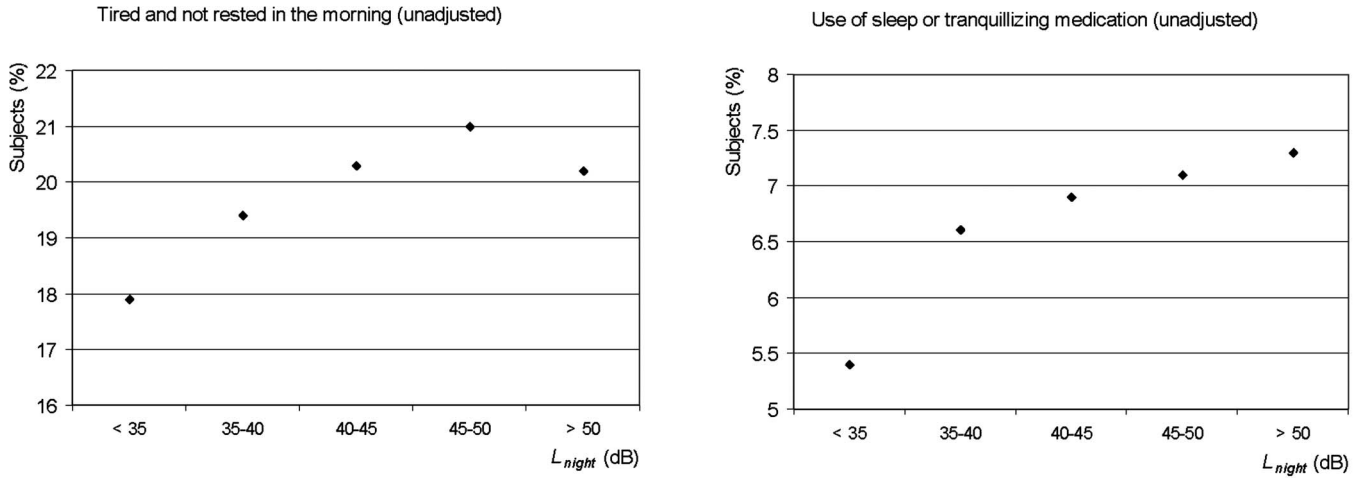


FIG. 2. Prevalence of self-reported sleep problems in subjects of the GLOBE study sample in relation to night time road traffic noise exposure at the home (L_{night}) (dB) unadjusted for confounders.

TABLE II. ORs for night time road traffic noise exposure (L_{night}) (dB) for markers of self-reported sleep problems (tired and not rested in the morning, use of sleep or tranquilizing medication) in the GLOBE study sample after adjustment for potential confounders. ORs from logistic regression are shown, adjusted for age, sex, BMI, exercise, marital status, work situation, financial difficulties, smoking, alcohol use, and education.

Group	N	OR $L_{night}<35$	OR $L_{night}35-40$	OR $L_{night}40-45$	OR $L_{night}45-50$	OR $L_{night}>50$
Tired and not rested in the morning	17 821	1.00	1.08 (0.95–1.22)	1.18 (1.03–1.34)[a]	1.26 (1.09–1.45)[b]	1.15 (1.00–1.33)[a]
Use of sleep or tranquilizing medication	17 855	1.00	1.14 (0.92–1.41)	1.16 (0.93–1.44)	1.21 (0.96–1.54)	1.15 (0.92–1.45)

[a]Significant relationship ($P<0.05$).
[b]Significant relationship ($P<0.01$).

Table III shows the ORs for covariates for markers of self-reported sleep problems (tired and not rested in the morning; use of sleep or tranquilizing medication) in the GLOBE study sample in the adjusted models for the association between night time road traffic noise exposure and sleep problems.

IV. DISCUSSION AND CONCLUSIONS

This study investigated the relationship between road traffic noise exposure during night time and two indicators of sleep problems: getting up tired and not rested in the morning and the use of sleep or tranquilizing medication. After adjustment for potential confounders, a significant association was found between road traffic noise exposure at the home and the risk of getting up tired and not rested in the morning. Although the prevalence of use of sleep or tranquilizing medication was higher at higher noise levels compared to the reference category ($L_{night}<35$ dB), after adjustment for potential confounders this association was not significant. Thus, no evidence was found of an effect of road traffic noise on sleep medication, although the findings do not contradict the significant increase in prevalence with increasing *aircraft* noise exposure during the late evening found by Franssen *et al.* (2004). The present finding on morning tiredness adds to the evidence from community surveys and field studies that long-term traffic noise is associated with self-reported sleep disturbance (Miedema and Vos, 2007) and may adversely affect self-reported sleep quality (Franssen and Kwekkeboom, 2003). Furthermore, the present results lend support to earlier indications from field studies that traffic noise may have after-effects the following day (Carter, 1996; Ouis, 1999; Öhrström, 1989; Öhrström and Skanberg, 2004). In addition, the results show that noise exposure plays a role among all possible causes of sleep problems, a conclusion that cannot be drawn on the basis of analyses with sleep disturbance questions explicitly referring to noise, as in Miedema and Vos (2007). Night time road traffic volume has previously been shown to be a risk factor for insomnia, and the prevalence of morning tiredness was increased in the insomniacs as compared to the non-insomniacs (Kageyama *et al.*, 1997). The description of the different severity criteria of insomnia (mild, moderate, and severe) includes "an almost nightly complaint of…not feeling rested after the habitual sleep episode" (The International Classification of Sleep Disorders). Since in our study night time road traffic noise exposure was found to be associated with morning tiredness (getting up tired and not rested in the morning), road traffic noise exposure may be hypothesized to induce or aggravate symptoms of insomnia.

A slight decrease in relative risk estimate was found for the highest noise exposure category. Similar to these findings, Öhrström *et al.* (2006a) found a decreasing effect of L_{night} on self-reported sleep quality in the highest noise category, which they attributed to the increased tendency to sleep with closed windows. In a field study by Griefahn *et al.* (2000), window closing behavior was the primary variable

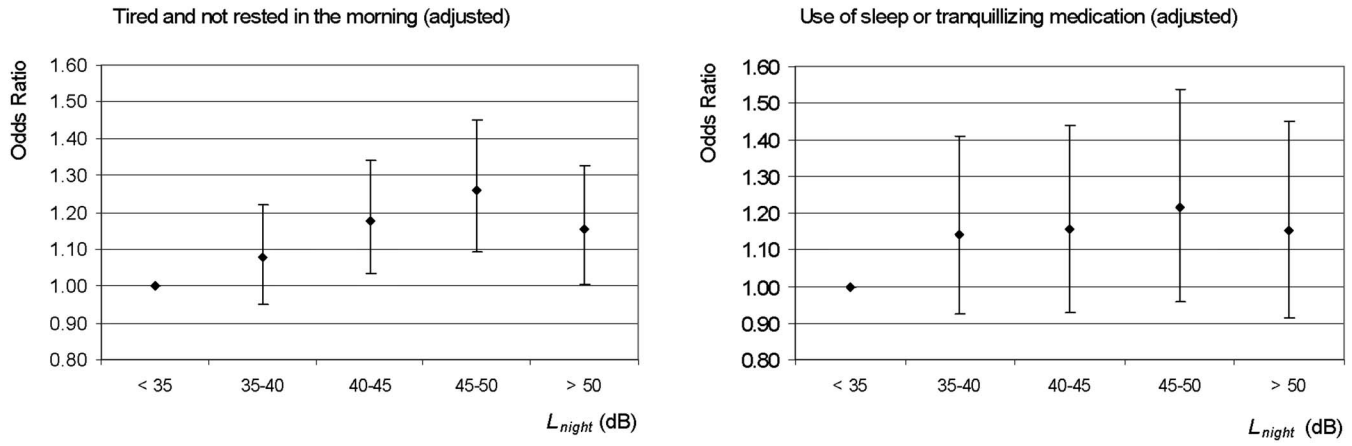


FIG. 3. ORs for night time road traffic noise exposure for markers of self-reported sleep problems (tired and not rested in the morning; use of sleep or tranquilizing medication) in the GLOBE study sample after adjustment for confounders.

TABLE III. ORs for covariates for markers of self-reported sleep problems (tired and not rested in the morning; use of sleep or tranquilizing medication) in the GLOBE study sample in the adjusted models for the association between night time road traffic noise exposure and sleep problems. ORs from logistic regression are shown.

	Tired, not rested in the morning	Sleep or tranquilizing medication use
Covariate	OR	OR
Age (year)	0.99[a]	1.04[a]
Sex (female)	1.23[a]	1.99[a]
BMI		
QI < 20	1.00	1.00
QI 20–25	0.95	0.72[a]
QI 25–30	0.96	0.70[a]
QI > 30	1.12	0.65[a]
Exercise		
None	1.00	1.00
Little	0.71[a]	0.87
Moderate	0.56[a]	0.61[a]
Much	0.45[a]	0.55[a]
Marital status		
Married	1.00	1.00
Unmarried	1.14[b]	0.98
Divorced	1.41[a]	1.88[a]
Widow(er)	1.10	1.38[a]
Work situation		
Unemployed	1.00	1.00
Working/studying	0.56[a]	0.38[a]
Other	0.57[a]	0.40[a]
Financial		
No difficulty	1.00	1.00
Some difficulty	1.43[a]	1.27[a]
Much difficulty	2.01[a]	1.56[a]
Smoking		
Never	1.00	1.00
Former	1.09	1.31[a]
Current	1.36[a]	1.73[a]
Alcohol		
Moderate	1.00	1.00
Abstainer	1.25[a]	1.60[a]
Excessive	1.05	1.16
Education[c]		
Category 1	1.00	1.00
Category 2	1.00	1.25
Category 3	0.98	1.32[b]
Category 4	1.19[b]	1.65[a]

[a]Significant relationship ($P < 0.01$).
[b]Significant relationship ($P < 0.05$).
[c]Highest attained level of education was distinguished into four different categories (higher professional education and university, intermediate professional and higher general education, lower professional and intermediate general education, and primary education).

associated with noise levels outside. A survey by Öhrström *et al.* (2006b) showed that L_{night} reduced both sleep quality and sleeping with open window. Another factor that may explain this decrease in effect is the self-selection of people less bothered by noise, particularly in areas with very high exposures (noise sensitive subjects moving away from high exposure areas). Effects of high noise exposure may be partly masked by this selection mechanism. In addition, better sound insulation of the dwellings of most exposed subjects and choice of bedroom location away from the source in reaction to exposure may affect relative risk estimates, particularly in the highest noise exposure category. Unfortunately, no data were available on noise sensitivity, dwelling insulation, choice of bedroom location, or window opening behavior to take these factors into account.

A limitation of the exposure assessment in this study is that road traffic noise data for the current situation (2004) were used. Unfortunately, no historic data for 1991 were available on traffic intensity and road characteristics such as road surface type and noise screens. Although traffic intensities may have increased, the road network is assumed to be rather stable, with only small (if any) changes in noise exposure across the population. For example, a recent study showed that correlations between road traffic intensities for a ten year period (1986–1996) in The Netherlands were high (>0.9) (Beelen *et al.*, 2009). If there is any effect of this limitation, it may be assumed that the actual association may be slightly stronger than found.

No data were available on noise exposures inside the bedrooms of the respondents. L_{night} at the most exposed façade as metric characterizes the exposure on one side of the dwelling, while the subject may sleep on another side of the dwelling, which may be less exposed. Having access to a quiet side of the dwelling has been found to reduce self-reported sleep disturbance and tiredness (Bluhm *et al.*, 2004; Öhrström *et al.*, 2006b). In addition to window opening behavior of subjects, the difference between this outside exposure level and the level inside the bedroom depends on the insulation of the façade. However, the tendency of people to sleep with their windows open is expected to reduce the variability in the outdoor-indoor difference.

Data for measures of sleep problems were available as dichotomous variables (response format: "yes" or "no"). Refined measurement of the effects might have refined the findings. If there would be an effect, it may be assumed that the association found in this study may be stronger still.

Strengths of the study include a number of aspects. First, the investigation was carried out in a large random sample drawn from the general population. The large sample size increases the power of the statistical analysis, while the population based design of the study increases the possibility to extrapolate the results to the general population, as compared to studies with smaller or non-random samples. In addition, the study was carried out for a large region, including Eindhoven City, which was the fifth largest city of The Netherlands at the start of the cohort study. As a result, there was a large variety in road traffic noise exposure, which may be expected to be representative for urbanized areas in general. The exposure was assessed with detailed noise models that take into account the relevant small scale intra-urban spatial variation in the study area. This approach reduces misclassification errors of noise exposure, which may occur in studies where exposure is based on subjective information (e.g., questionnaire reporting on traffic density or annoyance). Another strong point of this study was that it was not directed to

noise and its effects; thus no bias was introduced by subjects being triggered to focus on road traffic noise exposure. Finally, the authors were able to minimize confounding by adjusting for a large range of potential risk factors in the model, including age, sex, BMI, exercise, marital status, work situation, financial difficulties, smoking, alcohol use, and education. A sensitivity analysis was carried out to explore the effect of the inclusion of additional variables. Additional adjustment for major life events, number of children living at home, and cold or draught and dampness inside the dwelling did not substantially affect the association between night time road traffic noise and morning tiredness.

In conclusion, our results show that road traffic noise during the night is associated with after-effects: an increased risk of subjects getting up tired and not rested in the morning. These findings add to the evidence that residential road traffic noise exposure may cause sleep disturbance and could be interpreted as a signal that noise-induced sleep disturbance has significant implications for daily life in the general population. Furthermore, noise exposure may induce or aggravate symptoms of insomnia. It therefore appears to be important to increase awareness of transportation noise as a factor affecting sleep. Reduction of these effects may require specialized advice, for example, with respect to choice of bedroom location or measures of improving the sound insulation of the bedroom.

ACKNOWLEDGMENTS

The work reported in this article has been made possible by financial support of VROM. The GLOBE study is supported by grants from The Netherlands Ministry of Public Health, Welfare, and Sports and The Netherlands Prevention Fund. The study is conducted in close collaboration with the public Health Services of The Netherlands City of Eindhoven and the region of South-East Brabant. We thank the Municipality of Eindhoven for kindly providing data.

Aasvang, G. M., Moun, T., and Engdahl, B. (**2008**). "Self-reported sleep disturbances due to railway noise: Exposure-response relationships for nighttime equivalent and maximum noise levels," J. Acoust. Soc. Am. **124**, 257–268.

Basner, M., Samel, A., and Isermann, U. (**2006**). "Aircraft noise effects on sleep: Application of the results of a large polysomnographic field study," J. Acoust. Soc. Am. **119**, 2772–2784.

Beelen, R., Hoek, G., Houthuijs, D., Van den Brandt, P., Goldbohm, R. A., Fischer, P., Schouten, L. J., Armstrong, B., and Brunekreef, B. (**2009**). "The joint association of air pollution and noise from road traffic with cardiovascular mortality in a cohort study," Occup. Environ. Med. **66**, 243–250.

Berglund, B., Lindvall, T., and Schwela, D. H. (**1999**). "Guidelines for Community Noise," World Health Organization: Protection of the Human Environment.

Bluhm, G., Nordling, E., and Berglind, N. (**2004**). "Road traffic noise and annoyance—An increasing environmental health problem," Noise Health **6**, 43–49.

Borst, H. C., and Miedema, H. M. E. (**2005**). "Comparison of noise impact indicators, calculated on the basis of noise maps of DENL," Acta Acust. **91**, 378–385.

Carter, N. L. (**1996**). "Transportation noise, sleep, and possible after-effects," Environ. Int. **22**, 105–116.

Carter, N. L., Hunyor, S. N., Crawford, G., Kelly, D., and Smith, A. J. (**1994**). "Environmental noise and sleep—A study of arousals, cardiac arrhythmia and urinary catecholamines," Sleep **17**, 298–307.

Di Nisi, J., Muzet, A., Ehrhart, J., and Libert, J. P. (**1990**). "Comparison of cardiovascular responses to noise during waking and sleeping in humans," Sleep **13**, 108–120.

Drosopoulos, S., Schulze, C., Fischer, S., and Born, J. (**2007**). "Sleep's function in the spontaneous recovery and consolidation of memories," J. Exp. Psychol. Gen. **136**, 169–183.

Elmenhorst, E. M., Elmenhorst, D., Luks, N., Maass, H., Vejvoda, M., and Samel, A. (**2008**). "Partial sleep deprivation: Impact on the architecture and quality of sleep," Sleep Med. **9**, 840–850.

Franssen, E. A. M. and Kwekkeboom, J. M. I. (**2003**). "Effecten van geluid door wegverkeer op de slap: Een systematische review van studies in de woonomgeving. (Effects of road traffic noise on sleep: A systematic review of studies in de residential environment)," Report No. 715120010/2003, RIVM, Bilthoven.

Franssen, E. A. M., van Wiechen, C. M. A. G., Nagelkerke, N. J. D., and Lebret, E. (**2004**). "Aircraft noise around a large international airport and its impact on general health and medication use," Occup. Environ. Med. **61**, 405–413.

Graham, J. M. A., Janssen, S. A., Vos, H., and Miedema, H. M. E. (**2009**). "Habitual traffic noise at home reduces cardiac parasympathetic tone during sleep," Int. J. Psychophysiol. **72**, 179–186.

Griefahn, B., Bröde, P., Marks, A., and Basner, M. (**2008**). "Autonomic arousals related to traffic noise during sleep," Sleep **31**, 569–577.

Griefahn, B., Schuemer-Kohrs, A., Schuemer, R., Moehler, U., and Mehnert, P. (**2000**). "Physiological, subjective and behavioural responses during sleep to noise from rail and road traffic," Noise Health **3**, 59–71.

Harvey, A. G., Stinson, K., Whitaker, K. L., Moskovitz, D., and Virk, H. (**2008**). "The subjective meaning of sleep quality: A comparison of individuals with and without insomnia," Sleep **31**, 383–393.

HCN (Health Council of The Netherlands) (**2004**). "The influence of night time noise on sleep and health," Report No. 2004/14E, HCN, Den Haag.

Hofman, W. F., Kumar, A., and Tulen, J. H. M. (**1995**). "Cardiac reactivity to traffic noise during sleep in man," J. Sound Vib. **179**, 577–589.

Holmes, A. L., Burgess, H. J., and Dawson, D. (**2002**). "Effects of sleep pressure on endogenous cardiac autonomic activity and body temperature," J. Appl. Physiol. **92**, 2578–2584.

Horne, J. A., Pankhurst, F. L., Reyner, L. A., Hume, K., and Diamond, I. D. (**1994**). "A field study of sleep disturbance: Effects of aircraft noise and other factors on 5,742 nights of actimetrically monitored sleep in a large subject sample," Sleep **17**, 146–159.

International Standards Organization (**2002**). "Acoustics—Description, measurement and assessment of environmental noise: Part 2. Determination of environmental noise levels," Report No. ISO/CD1996-2, Geneva, Switzerland

Kageyama, T., Kabuto, M., Nitta, H., Kurokawa, Y., Taira, K., Suzuki, S., and Takemoto, T. (**1997**). "A population study on risk factors for insomnia among adult Japanese women: A possible effect of road traffic volume," Sleep **20**, 963–971.

Langdon, F. J., and Buller, I. B. (**1977**). "Road traffic noise and disturbance to sleep," J. Sound Vib. **50**, 13–28.

Mackenbach, J. P., van de Mheen, H., and Stronks, K. (**1994**). "A prospective cohort study investigating the explanation of socioeconomic inequalities in health in The Netherlands," Soc. Sci. Med. **38**, 299–308.

Marks, A. and Griefahn, B. (**2007**). "Associations between noise sensitivity and sleep, subjectively evaluated sleep quality, annoyance, and performance after exposure to nocturnal traffic noise," Noise Health **9**, 1–7.

Marshall, L., and Born, J. (**2007**). "The contribution of sleep to hippocampus-dependent memory consolidation," Trends Cogn. Sci. **11**, 442–450.

Michaud, D. S., Fidell, S., Pearsons, K., Campbell, K. C., and Keith, S. E. (**2007**). "Review of field studies of aircraft noise-induced sleep disturbance," J. Acoust. Soc. Am. **121**, 32–41.

Miedema, H. M. E., and Vos, H. (**2007**). "Associations between self-reported sleep disturbance and environmental noise based on reanalyses of pooled data from 24 studies," Behav. Sleep Med. **5**, 1–20.

Murphy, T. I., Richard, M., Masaki, H., and Segalowitz, S. J. (**2006**). "The effect of sleepiness on performance monitoring: I know what I am doing, but do I care?" J. Sleep Res. **15**, 15–21.

Muzet, A. (**2007**). "Environmental noise, sleep and health," Sleep Med. Rev. **11**, 135–142.

Ohayon, M. D. (**2005**). "Prevalence and correlates of nonrestorative sleep complaints," Arch. Intern. Med. **165**, 35–41.

Öhrström, E. (**1989**). "Sleep disturbance, psycho-social and medical symptoms—A pilot survey among persons exposed to high levels of road traffic noise," J. Sound Vib. **133**, 117–128.

Öhrström, E. (**2000**). "Sleep disturbances caused by road traffic noise—Studies in laboratory and field," Noise Health **2**, 71–78.
Öhrström, E., Hadzibajramociv, E., Holmes, M., and Svensson, H. (**2006a**). "Effects of road traffic noise on sleep: Studies on children and adults," J. Environ. Psychol. **26**, 116–126.
Öhrström, E., and Skanberg, A. (**2004**). "Sleep disturbances from road traffic and ventilation noise—Laboratory and field experiments," J. Sound Vib. **271**, 279–296.
Öhrström, E., Skanberg, A., Svensson, H., and Gidlöf-Gunnarsson, A. (**2006b**). "Effects of road traffic noise and the benefit of access to quietness," J. Sound Vib. **295**, 40–59.
Ouis, D. (**1999**). "Exposure to nocturnal road traffic noise: Sleep disturbance and its after effects," Noise Health **4**, 11–36.
Passchier-Vermeer, W., and Passchier, W. F. (**2000**). "Noise exposure and public health," Environ. Health Perspect. **108**, 123–131.
Passchier-Vermeer, W., Vos, H., Janssen, S. A., and Miedema, H. M. E. (**2007**). "Slaap en verkeersgeluid (Sleep and traffic noise)," TNO Report No. 2007-D-R0012/A, TNO, Delft, The Netherlands.
Pearsons, K. S., Barber, D. S., Tabachnick, B. G., and Fidell, S. (**1995**). "Predicting noise-induced sleep disturbance," J. Acoust. Soc. Am. **97**, 331–338.
Raidy, D. J., and Scharff, L. F. (**2005**). "Effects of sleep deprivation on auditory and visual task," Percept. Mot. Skills **101**, 451–467.
Scott, L. D., Hwang, W. T., Rogers, A. E., Nysse, T., Dean, G. E., and Dinges, D. F. (**2007**). "The relationship between nurse work schedules, sleep duration and drowsy driving," Sleep **30**, 1801–1807.
Siegel, J. M. (**2005**). "Clues to the functions of mammalian sleep," Nature (London) **437**, 1264–1271.
Spiegel, K., Leproult, R., and Van Cauter, E. (**1999**). "Impact of sleep debt on metabolic and endocrine function," Lancet **354**, 1435–1439.
Thiessen, G. J., and Lapointe, A. C. (**1983**). "Effect of continuous traffic noise on percentage of deep sleep, waking and sleep latency," J. Acoust. Soc. Am. **73**, 225–229.
VROM (**2006**). "Meet- en rekenvoorschrift geluidhinder 2006 (Prescribed measurement and calculation methods for noise annoyance 2006)," Staatscourant **249**, 84.
Wright, C. E., Erblicht, J., Valdimarsdottir, H. B., and Bovbjerg, D. H. (**2007**). "Poor sleep the night before an experimental stressor predicts reduced NK cell mobilization and slowed recovery in healthy women," Brain Behav. Immun. **21**, 358–363.

Response to noise from modern wind farms in The Netherlands

Eja Pedersen[a)]
Halmstad University and University of Gothenburg, Halmstad University, P.O. Box 823, SE-301 18 Halmstad, Sweden

Frits van den Berg
University of Groningen and GGD Amsterdam, GGD Amsterdam, P.O. Box 2200, 1000 CE Amsterdam, The Netherlands

Roel Bakker
University Medical Centre Groningen, University of Groningen, A. Deusinglaan 1, 7913 AV Groningen, The Netherlands

Jelte Bouma
University Medical Centre Groningen, University of Groningen, A. Deusinglaan 1, 7913 AV Groningen, The Netherlands

(Received 3 December 2008; revised 14 April 2009; accepted 8 June 2009)

The increasing number and size of wind farms call for more data on human response to wind turbine noise, so that a generalized dose-response relationship can be modeled and possible adverse health effects avoided. This paper reports the results of a 2007 field study in The Netherlands with 725 respondents. A dose-response relationship between calculated *A*-weighted sound pressure levels and reported perception and annoyance was found. Wind turbine noise was more annoying than transportation noise or industrial noise at comparable levels, possibly due to specific sound properties such as a "swishing" quality, temporal variability, and lack of nighttime abatement. High turbine visibility enhances negative response, and having wind turbines visible from the dwelling significantly increased the risk of annoyance. Annoyance was strongly correlated with a negative attitude toward the visual impact of wind turbines on the landscape. The study further demonstrates that people who benefit economically from wind turbines have a significantly decreased risk of annoyance, despite exposure to similar sound levels. Response to wind turbine noise was similar to that found in Sweden so the dose-response relationship should be generalizable.
 [DOI: 10.1121/1.3160293]

PACS number(s): 43.50.Qp, 43.50.Rq, 43.50.Sr [BSF] Pages: 634–643

I. INTRODUCTION

Community noise is recognized as an environmental stressor, causing nuisance, decreased wellbeing, and possibly non-auditory adverse effects on health (Stansfeld and Matheson, 2003). The main sources of community noise are transportation and industry. Air transport is the most annoying of the dominant means of transport (Miedema and Oudshoorn, 2001), though at comparable sound levels noise from road traffic has the largest impact in terms of number of people affected. Increasing awareness of the adverse effects of noise has led to noise management recommendations, including guideline values to limit health effects in various situations (WHO, 2000) and action plans for reducing noise and preserving quietness (END, 2002), all with the aim of decreasing the overall noise load. Noise impact is quantified based on the relationship between noise dose and response, the latter measured as the proportion of the public annoyed or highly annoyed by noise from a specified source. Several studies have explored the community response to transportation noise. The results of all available studies have been synthesized and modeled to yield polynomials describing the expected proportion of people annoyed by road traffic, aircraft, or railroad noise (Miedema and Oudshoorn, 2001). Dose-response curves have also been modeled for noise from industry and shunting yards (Miedema and Vos, 2004), albeit in relatively few studies. The Lden (day–evening–night) noise exposure metric has been found to best describe the noise load from these sources (Miedema *et al.*, 2000). This metric is based on long-term equivalent sound pressure levels assessed for different times of the day, to which penalties of 5 dB for evening and 10 dB for nighttime hours are added. These penalties reflect the need for quietness at specific times of day when the background sound levels are assumed to be lower.

Wind turbines are a new source of community noise to which relatively few people have yet been exposed. The number of exposed people is growing, as in many countries the number of wind turbines is rapidly increasing. The need for guidelines for maximum exposure to wind turbine noise is urgent: While not unnecessarily curbing the development of new wind farms, it is also important to avoid possible adverse health effects. No generalized dose-response curves have yet been modeled for wind turbines, primarily due to the lack of results of published field studies. To the best of

[a)]Author to whom correspondence should be addressed. Electronic mail: eja.pedersen@hh.se

the authors' knowledge, there are only four such studies, all of which find different degrees of relationship between wind turbine sound levels and annoyance: (i) a European study carried out in Denmark, The Netherlands, and Germany (Wolsink *et al.*, 1993; Wolsink and Sprengers, 1993); (ii) a complementary Danish study (Pedersen and Nielsen, 1994); (iii) the first Swedish study (Pedersen and Persson Waye, 2004); and (iv) a more recent Swedish study (Pedersen and Persson Waye, 2007). The sizes and heights of wind turbines have increased over the time covered by these studies. The 1994 Danish study included 16 wind turbines of up to 150 kW nominal power with towers under 33 m high, while the latest Swedish study included wind turbines of up to 1.5 MW with towers up to 65 m high. Also, these studies included mostly single wind turbines, while groups of wind turbines, i.e., wind farms, are more common today. The studies all use the *A*-weighted equivalent sound pressure level under specific meteorological conditions (generally winds of 8 m/s at 10-m height and, implicitly, a neutral atmosphere) as the metric for the sound immission levels—the standard for describing the dose of wind turbine noise.

The results of these studies indicate that wind turbines differ in several respects from other sources of community noise. Modern wind turbines mainly emit noise from turbulence at the trailing edge of the rotor blades. The turbine sound power level varies with the wind speed at hub height. It also varies rhythmically and more rapidly as the sound is amplitude modulated with the rotation rate of the rotor blades, due to the variation in wind speed with height and the reduction in wind speed near the tower (Van den Berg, 2005, 2007). Amplitude-modulated sound is more easily perceived than is constant-level sound and has been found to be more annoying (Bradley, 1994; Bengtsson *et al.*, 2004). In addition, sound that occurs unpredictably and uncontrollably is more annoying than other sounds (Hatfield *et al.*, 2002; Geen and McCown, 1984).

Wind turbines are tall and highly visible, often being placed in open, rural areas with low levels of background sound and in what are perceived as natural surroundings. Consequently, wind turbines are sometimes regarded as visible and audible intruders in otherwise unspoiled environments (Pedersen *et al.*, 2007). Furthermore, the moving rotor blades draw attention, possibly enhancing the perception of sound in a multi-modal effect (Calvert, 2001).

In summary, wind turbine noise could be predicted to be easily perceived and—in some environments—annoying, depending on both sound levels and visual aspects. To assess possibly unacceptable adverse health effects, generalized dose-response relationships need to be estimated and related to those of other noise sources. To this end, a field study exploring the impact of wind turbine sound on people living in the vicinity of wind farms was carried out in The Netherlands in 2007. The objectives of this study, reported here, were (i) to assess the relationship between wind turbine sound levels at dwellings and the probability of noise annoyance, taking into account possible moderating factors, (ii) to explore the possibility of generalizing a dose-response relationship for wind turbine noise by comparing the results of this study with those of previous Swedish studies, and (iii) to relate annoyance with wind turbine noise to annoyance with noise from other sources.

II. METHOD

A. Site selection

The site selection was intended to reflect contemporary wind turbine exposure conditions over a range of background sound levels. All areas in The Netherlands with at least two wind turbines of at least 500 kW within 500 m of each other and characterized by one of three clearly defined land-use types (i.e., built-up area, rural area with a main road, and rural area without a main road) were selected for the study. Sites dominated by industry or business were excluded, as these are not representative for residential areas and detailed examination showed that most of the nearest dwellings were not in the industrial areas but far from the wind turbines, thus adding to the already over-populated sound level classes in the study group. In The Netherlands, 1735 wind turbines were operating onshore in March 2006, 1056 of which were of 500 kW or greater nominal electric power. To rule out short-term effects, sites that had changed over the March 2006–March 2007 period (when the study started) were excluded.

B. Study population and sample

The study population consisted of approximately 70 000 adults living within 2.5 km of a wind turbine at the selected sites. The study sample was selected stepwise: (i) The authors identified 4570 postal codes for the selected sites; (ii) for these postcodes, they obtained 17 923 addresses with individual *x* and *y* coordinates from Adrescoördinatenbestand Nederland (the Dutch coordinates file); (iii) these addresses were classified into 5-dB(A) intervals according to *A*-weighted sound immission level due only to wind turbine sound, i.e., <30, 30–35, 35–40, 40–45, and >45 dB(A); and (iv) further classified into the three area types. Statistical power calculations based on the results of previous studies indicated that approximately 150 respondents were required in each of the five immission level groups. As relatively few people were classified as belonging to the highest immission level groups, all people in these groups were assigned to the study sample. In the other groups, the sample was randomly selected, based on an expected response rate of 33%. The final study sample included 1948 people.

C. Assessments of immission levels

A-weighted sound power levels in octave bands (at 8 m/s wind speed at 10-m height in a neutral atmosphere) for all wind turbines ($n=1846$) at the selected sites were obtained from reports from consultancies, manufacturers, and reports used by local authorities, describing the results from sound power level measurements and used as input for calculating sound levels caused by wind turbines. When data were unavailable, which was more often the case for older and smaller wind turbines, the sound power level of a turbine of the same dimensions and electrical output was used. The

TABLE I. Study sample, number of respondents, and response rate according to 5-dB(A) sound level interval.

	Predicted *A*-weighted sound pressure levels [dB(A)]					
	<30	30–35	35–40	40–45	>45	Total
Study sample	473	494	502	282	197	1948
No. of respondents	185	219	162	94	65	725
Response rate (%)	39	44	32	33	33	37

propagation of sound from the wind turbines toward the dwellings of members of the study population was calculated in accordance with the model legally required in The Netherlands (VROM, 1999), the New Zealand standard as an example of a simple model (NZS, 1998), and the international ISO standard model (ISO, 1996). For all sites, the ground absorption was set to 1 (100% sound absorbing surface) and the receiver height to 5 m. *A*-weighted sound pressure levels of all wind turbines (including those of <500 kW nominal power) at the dwelling facade were added logarithmically. The values calculated in accordance with ISO will be used as the exposure variable in this paper.

D. Social survey

Subjective responses were obtained through a postal questionnaire presented as a survey investigating general living conditions but also including a section on road traffic and wind turbine noise. The questionnaire was based on one previously used in Swedish studies (Pedersen and Persson Waye, 2004, 2007). Response to wind turbine noise was measured using five different questions, all of which displayed high internal consistency (Cronbach's alpha=0.87). In the present study, response to wind turbine noise was based on the answer to the following question: "Below are a number of items that you may notice or that could annoy you when you spend time outdoors at your dwelling. Could you indicate whether you have noticed these or whether these annoy you?" This question was followed by a list of possible annoyance factors (i.e., olfactory, aural, or visual annoyances from different sources) of which wind turbine sound was one. The question could be answered on a five-point verbal rating scale, where 1="do not notice," 2="notice but not annoyed," 3="slightly annoyed," 4="rather annoyed," and 5="very annoyed." The question was repeated for indoor perception. The scale was dichotomized into "do not notice" (scale point 1) and "notice" (scale points 2–5) when perception was analyzed, into "not annoyed" (scale points 1–3) and "annoyed" (scale points 4–5) for analyzes of annoyance, and into "not very annoyed" (scale points 1–4) and "very annoyed" (scale point 5) for studies of highly annoyed. The whole five-pointed scale was used when correlations between perception and annoyance on one hand, and other variables such as sound pressure levels on the other were studied. Noise sensitivity was measured on a five-point scale ranging from "not at all sensitive" to "very sensitive." Attitudes toward the noise source were measured as the general opinion on wind turbines (general attitude) and on the visual impact of wind turbines on the landscape (visual attitude), as well as with eight polarized items, such as "pretty–ugly" and "dangerous–harmless," all on five-point scales. The questionnaire also contained questions regarding the possibility of hearing the sound under different meteorological conditions, how often the sound was regarded as annoying, whether wind turbines were visible from the dwelling, and whether the respondent benefited economically from the wind farm, either by full or partial turbine ownership, or by being compensated otherwise. The respondents were also asked to choose descriptors of the wind turbine noise from among several alternatives; these descriptors were partly derived from a previous experimental study of the perception of wind turbine sounds (Persson Waye and Öhrström, 2002).

Of the selected study sample, 37% satisfactorily completed and returned the questionnaire (Table I). There was no difference in immission levels between the respondents and non-respondents ($t=-0.38, p=0.703$). A random sample of non-responders ($n=200$) received a short questionnaire comprising only two of the questions from the original questionnaire, questions asking them to rate their annoyance with wind turbine noise outdoors and indoors on a scale of 0–10. There was no statistically significant difference in the answers to these two questions between the responders and followed-up non-responders ($t=-0.82, p=0.412$; $t=-0.74, p=0.458$).

E. Analysis methods

Response to wind turbine noise is presented as proportions of the number of respondents in each 5-dB(A) interval, 95% confidence intervals being calculated in accordance with Wilson (Altman *et al.*, 2000). Differences between proportions were tested using the Mann–Whitney U-test. Correlations between two variables with ordinal scales were explored using Spearman's rank correlation (r_s). Logistic regression analysis was used for the multivariate analyses, with a dichotomous response variable and a continuous scale of the exposure variable (*A*-weighted sound pressure level). The Hosmer–Lemeshow test was used to determine the fit of the regression models to the data; here a p-value >0.05 indicates a good fit, as no difference between modeled and observed data is desirable. Principal component analysis with Varimax rotation was used in constructing factors. All tests were two-sided and a p-value <0.05 was assumed to indicate statistical significance.

TABLE II. Response to wind turbine noise outdoors or indoors, proportion of respondents ($n=708$) according to 5-dB(A) sound level intervals, and 95% confidence intervals (95%CI).

	Predicted *A*-weighted sound pressure levels [dB(A)]				
	<30	30–35	35–40	40–45	>45
Outdoors, *n*	178	213	159	93	65
Do not notice (%) (95%CI)	75 (68–81)	46 (40–53)	21(16–28)	13 (8–21)	8 (3–17)
Notice, but not annoyed (%) (95%CI)	20 (15–27)	36 (30–43)	41 (34–49)	46 (36–56)	58 (46–70)
Slightly annoyed (%) (95%CI)	2 (1–6)	10 (7–15)	20 (15–27)	23 (15–32)	22 (13–33)
Rather annoyed (%) (95%CI)	1 (0–4)	6 (4–10)	12 (8–18)	6 (3–13)	6 (2–15)
Very annoyed (%) (95%CI)	1 (0–4)	1 (0–4)	6 (3–10)	12 (7–20)	6 (2–15)
Indoors, *n*	178	203	159	94	65
Do not notice (%) (95%CI)	87 (81–91)	73 (67–79)	61 (53–68)	37 (28–47)	46 (35–58)
Notice, but not annoyed (%) (95%CI)	11 (7–17)	15 (11–20)	22 (16–29)	31 (22–31)	38 (28–51)
Slightly annoyed (%) (95%CI)	1 (0–4)	8 (5–12)	9 (6–15)	16 (10–25)	9 (4–19)
Rather annoyed (%) (95%CI)	0 (0–2)	3 (1–6)	4 (2–8)	6 (3–13)	5 (2–13)
Very annoyed (%) (95%CI)	1 (0–4)	1 (0–4)	4 (2–8)	10 (5–17)	2 (0–8)

III. RESULTS

A. Response to wind turbine sound

The degree of perception and annoyance increased with increasing sound level, for both outdoor ($r_s=0.50, n=708, p<0.001$) and indoor annoyance ($r_s=0.36, n=699, p<0.001$). The distribution of the response variables in relation to the sound level intervals is shown in Table II. In the 35–40-dB(A) sound level interval, 78% of respondents noticed sound outdoors from wind turbines, in the 40–45-dB(A) interval 87% noticed, and in the >45-dB(A) interval 92% noticed. As expected, the sound was not as frequently noticed indoors.

The loudness of the wind turbine sound was perceived differently under different meteorological conditions. Of the respondents, 69% reported that the sound was louder than average when the wind was blowing from the wind turbines toward the dwelling (downwind conditions), vs 5% who reported that it was less loud under those conditions. In addition, 67% reported that the sound was louder downwind when the wind was strong vs 18% who reported that it was less loud, and 40% thought the sound was louder at night while 22% thought it was less loud. The rest of the respondents reported that there was no difference between sound levels or that they did not know. "Swishing/lashing" was the most common descriptor of the wind turbine sound used by those who noticed the sound from their dwellings (75% of $n=335$), followed by "rustling" (25%), and "a low-frequency/low-pitch sound" (14%). Less than 10% reported "whistling/screeching," "thumping/throbbing," "resounding," "a pure tone," or "scratching/squeaking."

The proportion of respondents who were annoyed (rather or very) by the sound increased with increasing sound level up to 40–45 dB(A), after which it decreased. 18% were annoyed in the 35–40- and 40–45-dB(A) intervals, and 12%

TABLE III. Description of possible moderating variables in relation to 5-dB(A) sound level intervals: proportion of respondents ($n=725$) per sound level interval and 95% confidence intervals (95%CI) for subjective variables.

	Predicted *A*-weighted sound pressure levels [dB(A)]				
	<30 $n=185$	30–35 $n=219$	35–40 $n=162$	40–45 $n=94$	>45 $n=65$
Economic benefits (%)	2	3	10	34	67
Situational parameters					
Wind turbine visible (%)	35	60	90	89	100
Rural area (%)	36	30	46	43	52
Rural area with main road (%)	27	32	36	38	46
Built-up area (%)	37	38	17	19	2
Subjective variables					
Noise sensitive (%) (95%CI)	36 (29–43)	25 (19–31)	31 (24–38)	31 (22–41)	23 (15–35)
Negative attitude (%) (95%CI)	10 (7–16)	14 (10–19)	19 (13–25)	17 (11–26)	9 (4–19)
Negative visual attitude (%) (95%CI)	33 (26–40)	36 (30–43)	45 (37–52)	39 (30–49)	20 (12–41)

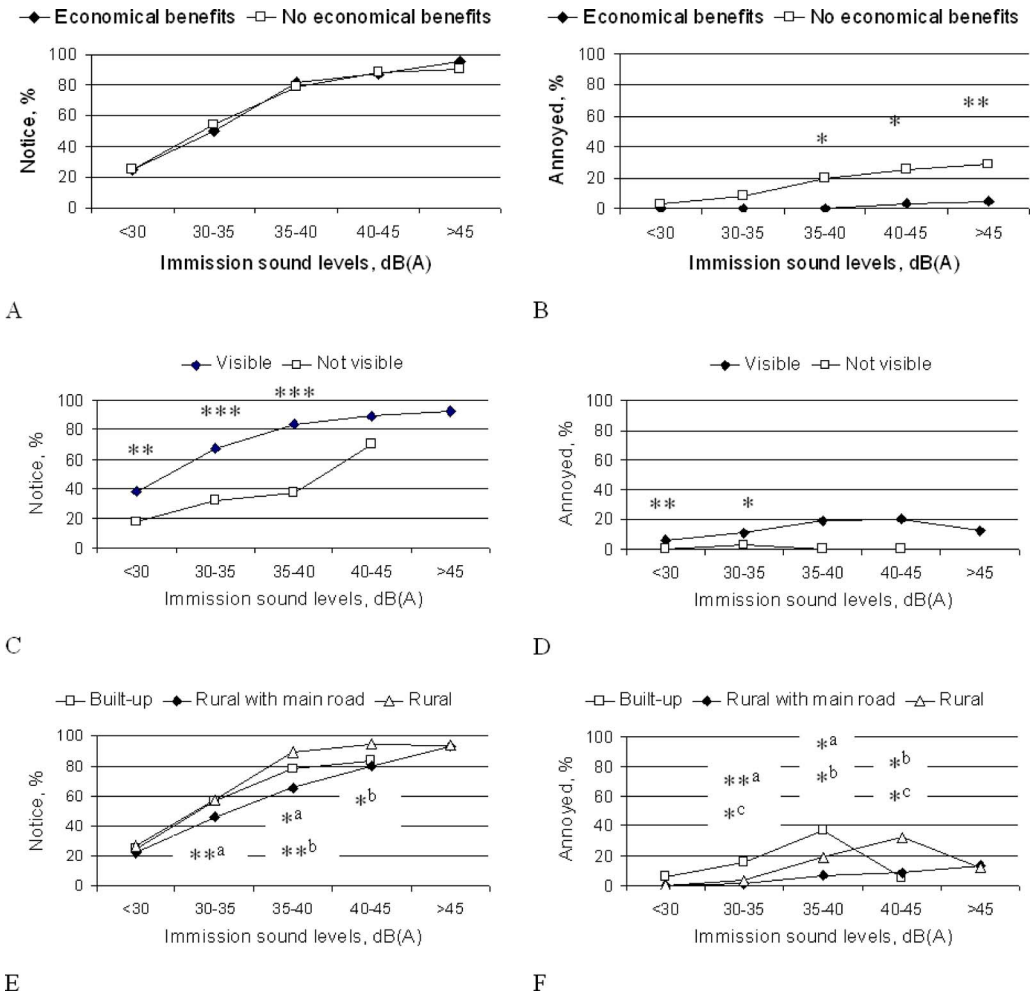


FIG. 1. (Color online) Proportion of respondents in each sound immission interval who noticed sound from wind turbines outside their dwelling (left column) or were annoyed by it (right column); [(A) and (B)] comparison between respondents who benefited economically and those who did not; [(C) and (D)] comparison between respondents who could see at least one wind turbine from their dwelling and those who could not; [(E)–(F)] comparison between respondents living in built-up areas, rural areas with a main road, and rural areas without a main road. $^{*}=p<0.05$, $^{**}=p<0.01$, $^{***}=p<0.001$ (Mann–Whitney U-test), ^{a}=built-up vs rural with main road, ^{b}=rural with main road vs rural without main road, and ^{c}=built-up vs rural without main road.

at levels above 45 dB(A) (Table II, outdoors). Almost all of the respondents that were annoyed by wind turbine sound had also reported that they were annoyed by sound from the rotor blades once a week or more often (92%). The proportions of respondents annoyed indoors were lower: In the sound level intervals below 40 dB(A) less than 10% were rather or very annoyed by the noise indoors, at 40–45 dB(A) 16% reported annoyance, and at levels above 45 dB(A) 6% were annoyed.

B. Moderating factors

Of the respondents, 100 reported that they benefited economically from the wind turbines, either by full or partial turbine ownership, or by receiving other economic benefits. Most of these respondents were subject to higher sound levels (Table III), 76 being subject to a level above 40 dB(A). There was no difference in terms of noticing wind turbine sound between those who benefited economically and those who did not [Fig. 1(A)], though there was a difference in *annoyance* [Fig. 1(B)]. Only 3 of the 100 respondents who benefited economically reported being annoyed by wind turbine sound.

Almost all respondents subject to sound pressure levels above 35 dB(A) could see at least one wind turbine from outside or inside their dwelling (Table III). The proportion of respondents who noticed sound from wind turbines [Fig. 1(C)], as well as the proportion annoyed by the noise [Fig. 1(D)], was larger for those who could see wind turbines from their dwellings than for those who could not. Only a few respondents who could not see any wind turbines were annoyed by the noise, even in the higher sound level intervals.

The distribution of respondents between the three types of area was fairly even in the lower sound level intervals, but at the higher sound intervals only a few people lived in built-up areas (Table III). Figure 1(E) indicates that at higher levels it was easier to notice wind turbine sound in rural areas without any main roads than it was in built-up areas, and that the sound was less noticeable in rural areas with a main road. In the lower sound level intervals, however, annoyance was more common in built-up than in both types of rural areas [Fig. 1(F)]. The proportions of respondents who benefited economically were higher in the two types of rural areas (19%) than in the built-up areas (2%). The wind turbines, on the other hand, were more visible in the rural areas,

TABLE IV. Results of two logistic regression models using the response variables *do not notice/notice* and *not annoyed/annoyed*, respectively; the exposure variable *sound pressure level* (continuous scale) and situational factors were used as moderating variables ($n=680$).

	Estimate (B)[a]	SE[b]	*p*-value	Exp(b)[c]
Do not notice vs notice (H–L)[d] ($p=0.721$)				
Sound pressure level [dB(A)]	0.17	0.022	<0.001	1.2
Economic benefits (no/yes)	−0.04	0.376	0.911	1.0
Visibility (no/yes)	1.40	0.214	<0.001	4.1
Area type (reference: rural)				
Rural with main road	−0.74	0.231	<0.01	0.5
Built-up	−0.18	0.240	0.451	0.8
Not annoyed vs annoyed (H–L)[d] ($p=0.199$)				
Sound pressure level [dB(A)]	0.13	0.027	<0.001	1.1
Economic benefits (no/yes)	−2.77	0.665	<0.001	0.1
Visibility (no/yes)	2.62	0.740	<0.001	13.7
Area type (reference: rural)				
Rural with main road	−1.07	0.372	<0.01	0.3
Built-up	0.65	0.321	<0.05	1.9

[a]Coefficients of the independent variables in the logistic regression.
[b]Standard errors of the coefficients.
[c]The exponential function of the coefficients of the independent variables in the logistic regression, which corresponds to the odds' ratio.
[d]Hosmer–Lemeshow goodness-of-fit test; *p*-value >0.05 indicates there is no statistically significant difference between the modeled and the observed data.

where 73% of respondents could see at least one wind turbine from their dwellings vs 54% in built-up areas.

The situational variables were entered simultaneously in logistic regressions. First, "do not notice wind turbine sound" vs "notice" was designated a dependent variable. The probability of hearing the sound was greater if the wind turbines were visible than if they were not. At the same time, living in a rural area with a main road as opposed to an area without decreased the probability (Table IV). Economic benefits had no statistically significant impact on perception of the sound. In the second regression model, the dichotomous variable "not annoyed by wind turbine sound" vs "annoyed" was designated as dependent. As before, the probability of being annoyed by wind turbine sound was higher if wind turbines were visible than if they were not (Table IV). Respondents who benefited economically were less likely to be annoyed than those who did not benefit. Living in a built-up area as opposed to a rural area without a main road increased the probability of being annoyed, while living in a rural area with a main road decreased the probability. Both regression models displayed good fit.

Approximately one of three respondents reported being rather or very sensitive to noise (Table III). There was no statistically significant relationship between noise sensitivity and the sound pressure level of wind turbine noise; there was, however, a positive correlation between noise sensitivity and annoyance (Table V). Noise sensitivity was also correlated with attitude toward the noise source. Of the respondents, 14% were negative (rather or very) toward wind turbines in general (general attitude), and 36% were negative toward the visual impact of wind turbines on the landscape (visual attitude). Attitude was not related to sound levels, but

TABLE V. Correlations between sound pressure levels, response (five-point scale from "do not notice" to "very annoyed"), and subjective variables; Spearman's rank correlation test.

	1	2	3	4
1. Sound pressure level [dB(A)]	⋯			
2. Response (five-point scale)	0.51[a]	⋯		
3. Noise sensitivity (five-point scale)	−0.01	0.14[a]	⋯	
4. General attitude (five-point scale)	−0.03	0.24[a]	0.14[a]	⋯
5. Visual attitude (five-point scale)	−0.01	0.29[a]	0.26[a]	0.65[a]

[a]$p<0.001$.

to annoyance (Table V). The association between noise annoyance on the one hand, and the variables noise sensitivity, general attitude, and visual attitude, on the other, was confirmed by testing in a logistic regression with the dependent variable "not annoyed" vs "annoyed" and adjusting for sound levels (Table VI). Of the three variables, visual attitude (i.e., attitude toward the visual impact of wind turbines on the landscape) had the strongest relationship with annoyance.

According to the eight polarized items, the wind turbines on average tended to be rated as relatively ugly (vs pretty), repulsive (vs inviting), unnatural (vs natural), and annoying (vs blending in); also, they were rated as efficient (vs inefficient), environmentally friendly (vs not environmentally friendly), necessary (vs unnecessary), and harmless (vs dangerous). Principal component analysis revealed that six of these items could be grouped to form two constructed factors: (i) *visual judgments*, comprising "pretty–ugly," "inviting–repulsive," and "natural–unnatural" (Cronbach's alpha=0.850), and (ii) *utility judgments*, comprising "environmentally friendly–not environmentally friendly," "efficient–inefficient," and "necessary–unnecessary" (Cronbach's alpha=0.804). These two factors accounted for 75% of the variance of the included items. The factor visual judgments was highly correlated with visual attitude ($r_s=0.602, p<0.001$) and, to a lesser degree, with general attitude toward wind turbines ($r_s=0.501, p<0.001$). The factor utility judgments was more highly correlated with general attitude ($r_s=0.513, p<0.001$) than with visual attitude ($r_s=0.381, p<0.001$).

TABLE VI. Results of a logistic regression model with the response variables *not annoyed/annoyed*, the exposure variable *sound pressure level* (continuous scale), and individual factors as moderating variables ($n=670$).

	Estimate (B)[a]	SE[b]	*p*-value	Exp(b)[c]
Not annoyed vs annoyed (H–L)[d] ($p=0.977$)				
Sound pressure level [dB(A)]	0.10	0.025	<0.001	1.1
Noise sensitivity (five-point scale)	0.35	0.138	<0.05	1.4
General attitude (five-point scale)	0.54	0.172	<0.01	1.7
Visual attitude (five-point scale)	1.04	0.215	<0.001	2.8

[a]Coefficients of the independent variables in the logistic regression.
[b]Standard errors of the coefficients.
[c]The exponential function of the coefficients of the independent variables in the logistic regression, which corresponds to the odds' ratio.
[d]Hosmer–Lemeshow goodness-of-fit test; *p*-value >0.05 indicates there is no statistically significant difference between the modeled and the observed data.

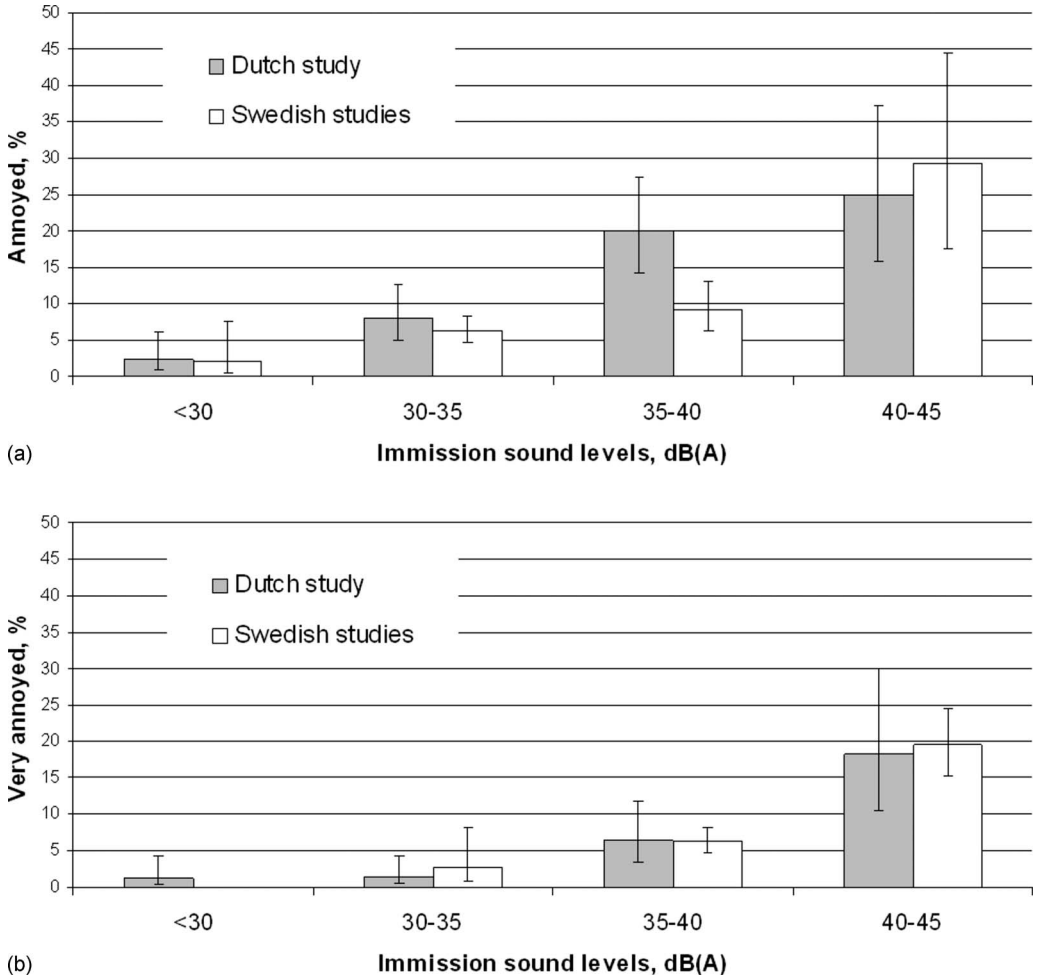


FIG. 2. Proportions of respondents annoyed (a) and very annoyed (b) by wind turbine noise outside their dwellings in four sound level intervals in the Dutch study (only respondents who did not benefit economically, $n=586$) and the Swedish studies ($n=1095$), with 95% confidence intervals.

C. Sound exposures

The calculated immission levels representing the levels of wind turbine sound outside respondent dwellings were similar for all three different calculation methods, and these values were highly correlated with each other ($r^2>0.98$). The differences between the levels calculated in accordance with the ISO standard and those calculated using the Dutch algorithm ranged from −0.8 to 1.4 dB(A), the average difference being 0.3 dB(A). The differences between the levels calculated in accordance with the ISO and the New Zealand standards were somewhat greater, −4.4 to 1.8 dB(A), the average difference being −0.8 dB(A). There were no differences in the exposure-response relationships between the three different methods for calculating the exposure levels; the correlation coefficient for the relationship between sound levels and response to wind turbine noise outdoors was 0.50 in all three cases.

IV. COMPARISONS WITH SWEDISH STUDIES

The proportions of respondents annoyed by wind turbine noise were compared with similar merged data from the two previous Swedish studies (Pedersen and Persson Waye, 2004, 2007) (Fig. 2). In the Swedish studies, the *A*-weighted sound levels were calculated in accordance with the Swedish standard, which uses a simplified algorithm comparable to the New Zealand standard for receiver points 1000 m or less from a wind turbine, and an algorithm based on octave bands for greater distances. The sound power used was the level at a wind speed of 8 m/s at 10-m height assuming a neutral atmosphere, as in the present study. The high degree of agreement between the dose calculations, as demonstrated above, leads to the assumption that calculation of the sound levels in the Swedish study was comparable to those used in the Dutch study. In the Swedish studies, many of the respondents (77% of 1059) could see at least one wind turbine from their dwellings, just as in the Dutch study. However, almost none of the Swedish respondents benefited economically from the turbines. The response to wind turbine noise found in the Swedish studies will therefore be compared with the responses of those not benefiting economically from the turbines in the Dutch study. Annoyance (rather or very) with wind turbine noise displays great agreement between the studies for the lowest sound level intervals [Fig. 2(A)]. For the 35–40-dB(A) interval, annoyance was greater among Dutch than Swedish respondents, the difference being statistically significant for respondents not benefiting economically from the turbines. In contrast, for the 40–45-dB(A) interval, the proportion of respondents annoyed was somewhat smaller in the Dutch study than in the Swedish studies,

though this difference was not significant. No Swedish data were available for higher sound levels. No differences between Dutch and Swedish respondents were found when comparing the percentages of very annoyed respondents [Fig. 2(B)].

V. COMPARISONS WITH OTHER SOURCES OF COMMUNITY NOISE

The proportions of respondents annoyed by wind turbine noise were compared with the proportions annoyed by other sources of community noise. For transportation, third-order polynomials with the Lden exposure metric have been formulated by Miedema and Oudshoorn (2001). These models are based on 19 aircraft studies, 26 road traffic studies, and 8 railway studies. The polynomials were forced to zero at Lden=37 dB(A) for moderate annoyance and to 42 dB(A) for severe annoyance, i.e., it was assumed that no moderate/severe annoyance with transportation noise occurred below these levels. For stationary sources, second-order polynomials were used, also with the Lden exposure metric (Miedema and Vos, 2004). The data originate from one study of eight industries (not seasonal) and two shunting yards.

The *A*-weighted sound pressure levels used as the exposure variable in the wind turbine studies must be converted into the Lden metric to enable comparisons. The sound power level of a wind farm changes with wind speed, so the long-term equivalent level depends on the wind speed distribution at the hub. Van den Berg (2008) suggested that *A*-weighted sound pressure levels capturing conditions at a wind of 8 m/s at 10-m height could be transformed into Lden values by adding 4.7 ± 1.5 dB. These findings are based on the long-term measurement of wind speed at hub height of modern wind turbines, also taking into account the atmospheric states (stable, neutral, or unstable) during the day and night and different locations (coastal and inland). As meteorological data from the wind farm sites used in the present study were unavailable, this simplified transformation was used for the exposure variable in this study, i.e., 4.7 dB were added to the calculated immission levels.

To allow comparisons between studies, Miedema and Vos (1998) suggested standardized transformations of the proportion of annoyed respondents measured at different scales. The base is a scale from 0 (no annoyance at all) to 100 (very annoyed). The cutoff point for the proportion of respondents annoyed is 50 and for highly annoyed 72. The scale used in the present study uses two scale points in reporting no annoyance: 1—"do not notice" and 2—"notice, but not annoyed." These two scale points were merged into one, to reduce the five-point scale to a four-point scale. Hence, with a cutoff point of 50 on a 0–100 scale, the proportion of respondents annoyed was represented by those reporting 4—"rather annoyed" and 5—"very annoyed." Similarly, respondents reporting 5—"very annoyed" could be compared to those highly annoyed in previous studies. Only respondents who did not benefit economically from wind turbines were included, as it can be assumed that few respondents benefited economically from these other noise sources. The curves describing the dose-response relationship for

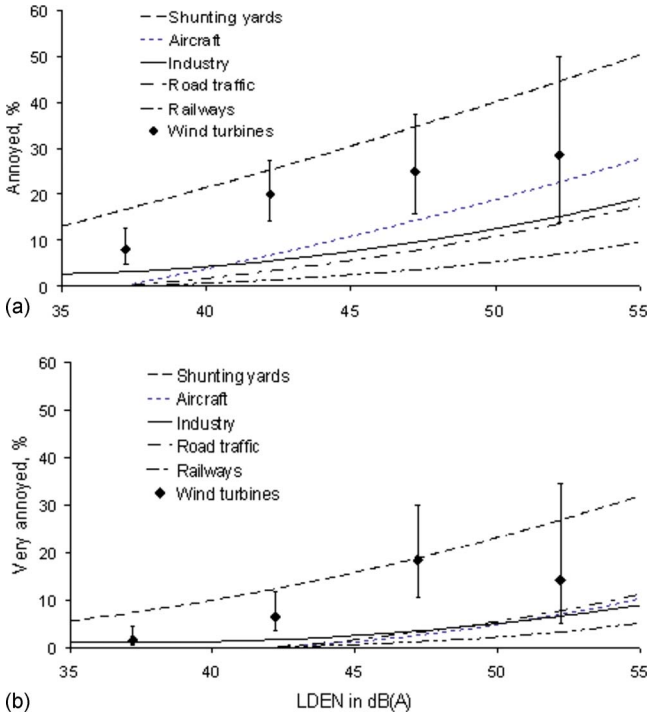


FIG. 3. (Color online) Proportion of respondents annoyed (a) and very annoyed (b) by wind turbine noise outside their dwellings (only respondents who did not benefit economically, $n=586$) compared to the modeled response [(A) percentage annoyed and (B) percentage highly annoyed] to noise from road traffic, aircraft, and railways (Miedema and Oudshoorn, 2001) and from industry and shunting yards (Miedema and Vos, 2004). For wind turbine noise the median of sound immission levels in each 5-dB interval is at the abscissa.

sound sources other than wind turbines do not distinguish between outdoor and indoor responses. Here, outdoor annoyance with wind turbine noise was chosen for comparison.

The comparison shows that the proportion of respondents annoyed with wind turbine noise below 50 dB(A) Lden is larger than the proportion annoyed with noise from all other noise sources except shunting yards (Fig. 3). At higher sound levels, this is less certain due to the low number of respondents leading to large confidence intervals.

VI. DISCUSSION

Noise from wind turbines was found to be more annoying than noise from several other sources at comparable Lden sound levels. The proportions of people annoyed by wind turbine noise lie between the proportions expected to be annoyed by noise from aircraft and from shunting yards. Like aircraft, wind turbines are elevated sound sources visible from afar and hence intrude both visually and aurally into private space (Brown, 1987). A strong correlation between noise annoyance and negative opinion of the impact of wind turbines on the landscape was found in early studies of perceptions of wind turbines (Wolsink and Sprengers, 1993); this was confirmed in the present study, as manifested by words such as "ugly," "repulsive," and "unnatural." Three different landscapes were explored. Surprisingly, annoyance was highest in what was classified as built-up area, in this case, mostly small towns and villages. It cannot be excluded that reflections from buildings may have caused higher

sound levels than those calculated, though it is more plausible that nearby buildings would have reduced the noise. The higher annoyance levels found in towns could instead be interpreted as an effect of place attachment (Giuliani and Feldman, 1993). In this view, new technical devices being deemed not beneficial for the living environment induce a negative reaction (Lazarus and Cohen, 1977). This theory cannot, however, be confirmed from the present data set.

Previously, the relatively high annoyance with shunting yard noise has partly been explained by the impulsive nature of some yard activities (Miedema and Vos, 2004). Wind turbine sound also varies unpredictably in level within a relatively short time span, i.e., minutes to hours. It can be postulated that it could be even more important that neither type of noise ceases at night. In contrast, in areas with traffic noise and/or industrial noise, background levels usually return to lower levels at night, allowing residents to restore themselves psycho-physiologically. A large proportion of respondents in the present study reported hearing wind turbine sound more clearly at night, an observation supported by previous findings that, due to atmospheric conditions that are common over land in the temperate climate zone, nighttime immission levels can be higher than estimated from 10-m wind speeds using a neutral wind speed profile (Van den Berg, 2007) and also because the average hub height wind speed at night is higher than the same wind in day time (Van den Berg, 2008). In contrast, the near surface wind at night is often weaker in conditions that favor stronger high altitude winds, resulting in less wind-induced background sound from vegetation (Van den Berg, 2007) with less capacity to mask the wind turbine sound or even distract attention for it. Taken together, this implies that nighttime conditions should be treated as crucial in recommendations for wind turbine noise limits.

Using only the subsample that did not benefit economically from wind turbines in the comparisons with other noise sources could be questioned as the databases for the dose-response curves for other sources than wind turbines do not take (perceived) benefits into account. However, in contrast to the other noise sources (in Fig. 3), wind turbines are special because they can provide direct profit to residents in a wind farm area. They could be owned by a single person or by a group of people, or the landowner could receive a yearly income. The benefits from the other noise sources are not as direct and not as clearly economical, but they could be taken into account in future studies with the aim to compare noise annoyance due to different sources considering the benefits of each source.

There was great agreement as to how to describe the wind turbine sound. The dominant quality of the sound was swishing, a quality previously found to be the most annoying (Pedersen and Persson Waye, 2004). Few respondents described the turbine sound as low frequency, in line with recent reports confirming that modern wind turbines do not produce high levels of (audible) low-frequency sound (Jakobsen, 2005).

The proportion of annoyed respondents found by the present study was similar to that found by previous Swedish studies, indicating there were no cultural differences in the perception or appreciation of the sound between these two countries. However, annoyance was found to be significantly higher in the Dutch study in the 35–40-dB(A) interval. At these levels, it could be hypothesized that masking by background sound could have a large influence. The perceived difference could be due to the larger wind turbines included in the present study. Higher towers push the rotors to heights with stronger winds than found lower down, increasing the time a wind turbine operates and increasing differences between immission levels and the background sound levels of wind-induced noise in bushes and trees, especially at night when the atmosphere is stable for part of the time.

This study found a stronger relationship between immission levels of wind turbine noise and annoyance than the previously reported Swedish studies. This could be due to the study design, which, rather than concentrating on sampling participants from only a few areas, sampled participants from all suitable wind farm areas in The Netherlands, thus avoiding the influence of uncontrollable local factors. The nonacoustical factor that had the highest impact on noise annoyance was economic benefit, which substantially decreased the probability of annoyance. As was expected, people benefiting economically from a noise source are less likely to be annoyed by it, though to the best of the authors' knowledge this has not previously been demonstrated as clearly as in this study. The observed gap in annoyance between those benefiting economically and those who do not could be due to a more positive appraisal of the sound if it signifies profit. On the other hand, resentment against profiting neighbors among those not benefiting could have increased the annoyance in this group, also contributing to the gap. The study design with respondents from all over The Netherlands instead of fewer selected study areas should reduce the risk for local disputes to affect the results, unless these disputes, and the resentment they could cause, are occurring everywhere.

VII. CONCLUDING REMARKS

This study enlarges the basis for calculating a generalized dose-response curve for wind turbine noise usable for assessing wind turbine noise in terms of its environmental health impact, the number of people influenced by it, and, by extension, its role from a public health perspective. The study confirms that wind turbine sound is easily perceived and, compared with sound from other community sources, relatively annoying. Annoyance with wind turbine noise is related to a negative attitude toward the source and to noise sensitivity; in that respect it is similar to reactions to noise from other sources. This may be enhanced by the high visibility of the noise source, the swishing quality of the sound, its unpredictable occurrence, and the continuation of the sound at night. The study demonstrates that it is possible to model a highly needed generalized dose-response relationship for Northern Europe, and supposedly also for the rest of Europe and North America, if the different proportions of people benefiting economically from wind turbines in the different regions are taken into account. The study also

shows that mitigation measures can be directed to acoustical as well as non-acoustical factors that contribute to the impact of wind farms.

ACKNOWLEDGMENTS

The study was financed by the European Union as a Specific Support Action, Contract No. 044628. The authors also thank former students Jan Oudelaar, Eildert Slim, and Rowan Rossing for their help, and Wind Service Holland for providing the wind turbine location data.

Altman, D. G., Machin, D., Bryant, T. N., and Gardner, M. J. (**2000**). *Statistics With Confidence*, 2nd ed. (BMJ Books, London).

Bengtsson, J., Persson Waye, K., and Kjellberg, A. (**2004**). "Sound characteristics in low frequency noise and their relevance for the perception of pleasantness," Acta. Acust. Acust. **90**, 171–180.

Bradley, J. S. (**1994**). "Annoyance caused by constant-amplitude and amplitude-modulated sound containing rumble," Noise Control Eng. J. **42**, 203–208.

Brown, B. B. (**1987**). "Territoriality," in *Handbook of Environmental Psychology*, edited by D. Stokols and I. Altman (Wiley, New York).

Calvert, G. A. (**2001**). "Crossmodal processing in the human brain: Insights from functional neuroimaging studies," Cereb. Cortex **11**, 1110–1123.

END (**2002**). "European Noise Directive 2002/49/EC of the European Parliament and of the Council, of 25 June 2002, relating to the assessment and management of environmental noise," European Parliament and Council.

Geen, R. G., and McCown, E. J. (**1984**). "Effects of noise and attack on aggression and physiological arousal," Motiv. Emot. **8**, 231–241.

Giuliani, M. V., and Feldman, R. (**1993**). "Place attachment in a developmental and cultural context," J. Environ. Psychol. **13**, 267–274.

Hatfield, J., Job, R. F., Hede, A. J., Carter, N. L., Peploe, P., Taylor, R., and Morrell, S. (**2002**). "Human response to environmental noise: The role of perceived control," J. Behav. Med. **9**, 341–359.

ISO (**1996**)."Attenuation of sound during propagation outdoors. Part 2: General method of calculation," ISO 9613-2, International Organization for Standardization, Geneva.

Jakobsen, J. (**2005**). "Infrasound emission from wind turbines," Low Freq. Noise, Vib., Act. Control **24**, 145–155.

Lazarus, R. S., and Cohen, J. B. (**1977**). "Environmental stress," in *Human Behavior and Environment: Advances in Theory and Research*, edited by I. Altman and J. F. Wohlwill (Plenum, New York), Vol. **2**.

Miedema, H. M. E., and Oudshoorn, C. G. M. (**2001**). "Annoyance from transportation noise; relationships with exposure metrics DNL and DENL and their confidence intervals," Environ. Health Perspect. **109**, 409–416.

Miedema, H. M. E., and Vos, H. (**1998**). "Exposure-response relationships for transportation noise," J. Acoust. Soc. Am. **104**, 3432–3445.

Miedema, H. M. E., and Vos, H. (**2004**). "Noise annoyance from stationary sources: Relationships with exposure metric day-evening-night level (DENL) and their confidence intervals," J. Acoust. Soc. Am. **116**, 334–343.

Miedema, H. M. E., Vos, H., and de Jong, R. G. (**2000**). "Community reaction to aircraft noise; time-of-day penalty and tradeoff between levels of overflights," J. Acoust. Soc. Am. **107**, 3245–3253.

NZS (**1998**). "The assessment and measurement of sound from wind generators," New Zealand Standard 6808:1998, Standards Council of New Zealand, Wellington, New Zealand.

Pedersen, E., Hallberg, L. R.-M., and Persson Waye, K. (**2007**). "Living in the vicinity of wind turbines—A grounded theory study," Qual. Res. Psychol. **4**, 49–63.

Pedersen, E., and Persson Waye, K. (**2004**). "Perception and annoyance due to wind turbine noise: A dose–response relationship," J. Acoust. Soc. Am. **116**, 3460–3470.

Pedersen, E., and Persson Waye, K. (**2007**). "Wind turbine noise, annoyance and self-reported health and wellbeing in different living environments," Occup. Environ. Med. **64**, 480–486.

Pedersen, T. H., and Nielsen, K. S. (**1994**). "Genvirkning af støj fra vindmøller (Annoyance by noise from wind turbines)," Report No. 150, DELTA Acoustic and Vibration, Lydtekniske Institut, Copenhagen.

Persson Waye, K., and Öhrström, E. (**2002**). "Psycho-acoustic characters of relevance for annoyance of wind turbine noise," J. Sound Vib. **250**, 65–73.

Stansfeld, S. A., and Matheson, M. P. (**2003**). "Noise pollution: Non-auditory effects on health," Br. Med. Bull. **68**, 243–357.

Van den Berg, F. (**2005**). "The beat is getting stronger: The effect of atmospheric stability on low frequency modulated sound of wind turbines," Low Freq. Noise, Vib., Act. Control **24**, 1–24.

Van den Berg, F. (**2007**). "The sound of high winds: The effect of atmospheric stability on wind turbine sound and microphone noise," Doctoral thesis, Rijksuniversiteit, Groningen, The Netherlands.

Van den Berg, F. (**2008**). "Criteria for wind farm noise: Lmax and Lden," Proceedings of the Seventh European Conference on Noise control, EURONOISE, Acoustics "08, Paris, France.

VROM (**1999**). *Halndleiding Meten en Rekenen Insturielawaii (Manual for Measuring and Calculating Industrial Noise)* (Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer, den Haag, The Netherlands).

WHO (**2000**). *Guidelines for Community Noise*, edited by B. Berglund, T. Lindvall, D. H. Schwela, and K.-T. Goh (World Health Organization, Geneva).

Wolsink, M., and Sprengers, M. (**1993**). "Wind turbine noise: A new environmental threat?," Proceedings of the Sixth International Congress on the Biological Effects of Noise, ICBEN, Nice, France, Vol. **2**, pp. 235–238.

Wolsink, M., Sprengers, M., Keuper, A., Pedersen, T. H., and Westra, C. A. (**1993**). "Annoyance from wind turbine noise on sixteen sites in three countries," Proceedings of the European Community Wind Energy Conference, Lübeck, Travemünde, pp. 273–276.

A note on noise propagation in street canyons

Kai Ming Li
Ray W. Herrick Laboratories, School of Mechanical Engineering, Purdue University, 140 South Martin Jischke Drive, West Lafayette, Indiana 47907-2031

Chenly Yuen Cheung Lai
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

(Received 15 November 2008; revised 12 May 2009; accepted 13 May 2009)

The current study examines the propagation of sound in street canyons with geometrically reflecting surfaces. An image source method is a popular numerical model to estimate the propagation of sound energy in a street canyon. This numerical model calculates the total sound energy received at a field point by summing the contributions from individual image sources incoherently. The discrete image source model is generalized by replacing rows of point sources with their respective line sources. An integral formulation is derived, which can be evaluated exactly to give a simple analytical solution. The expression permits rapid computations of the sound energy due to a point source placed in a street canyon. The transient sound energy at a receiver point is also examined. It has been demonstrated that the transient sound energy can be expressed in terms of a standard exponential integral. The Schroeder integration method is then used to calculate the reverberation times, which allow a straightforward assessment of the acoustic environment in street canyons. Indoor and outdoor experiments were conducted to validate the proposed integral formulation. The analytical formulas were also compared with numerical results based on the standard image source method and with published experimental data. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3158599]

PACS number(s): 43.50.Rq, 43.50.Gf, 43.55.Dt, 43.55.Br [NX] Pages: 644–655

I. INTRODUCTION

In the past few decades, we witnessed a rapid growth in mechanized transport and transportation systems. Noise is one of the most cited environmental factors that are most commonly associated with pollution from transport. Various modes of land transportation are the primary source of noise in dense, high-rise cities. An essential feature of compact and high-rise cities is that the scarcity of suitable land has encouraged building development to go up in the vertical dimension. It is common to find residential and commercial tower blocks of over 40 storeys with height over 100 m or even more. Worse still, the urban areas are typically embedded with a compact traffic network of highways and/or railways with high traffic volume. The lack of available land space means that the residents are closer to noise sources. The tall flanking buildings, which form a street canyon, permit large portion of the dwellings exposing to the land transportation noise. The noise levels do not decrease significantly with the height above the ground in street canyons.[1–4] The motivation of the current study is to develop a simple formula to evaluate steady-state noise levels and the reverberation times in a street canyon for assessing the effectiveness of noise control.

It is of interest to point out that the first measurements of sound propagation in urban area were performed to determine an optimal location for sirens during 1940s.[5,6] Wiener *et al.*[7] conducted a theoretical study on the propagation of sound, the reverberation time, and speech intelligibility in a city street of Boston. In 1970s, Aylor *et al.*[8] conducted a study to investigate the effect of ivy grown on building façades on the reverberation time. Yeow[9,10] performed some measurements of the reverberation time in downtown residential areas. From their experimental results, Ko and Tang[11] suggested that the reverberation time of a street canyon was proportional to the volume bounded by the tall buildings in street canyons. Steenackers *et al.*[12] carried out a series of experimental measurements of the sound decay curves in a street canyon. Their study was focused on the determination of the absorption coefficient in favor of the reverberation time of the street canyons but no theoretical models were offered.

There were also a number of other studies addressing the theoretical and numerical aspects of sound propagation in street canyons in the past few decades. Typically, either an image source model or a diffuse reflection model is used to compute the sound fields and the reverberation times in a street canyon.[13–16] Most of the image source methods are based on incoherent model[17] although a coherent model has been used in a recent study.[18] Despite these theoretical and experimental studies, there is still no simple but accurate formula to predict the noise levels and the reverberation time in street canyons. The present study is aimed to develop such a simple but accurate analytical expression to predict the noise levels and to calculate the reverberation time rapidly. This, in turn, allows the assessment of the acoustic treatments in street canyons.

We remark that the scattering of sound at the façade surfaces is known to be significant for predicting the transmission of noise along a street canyon.[13,14] We shall use the

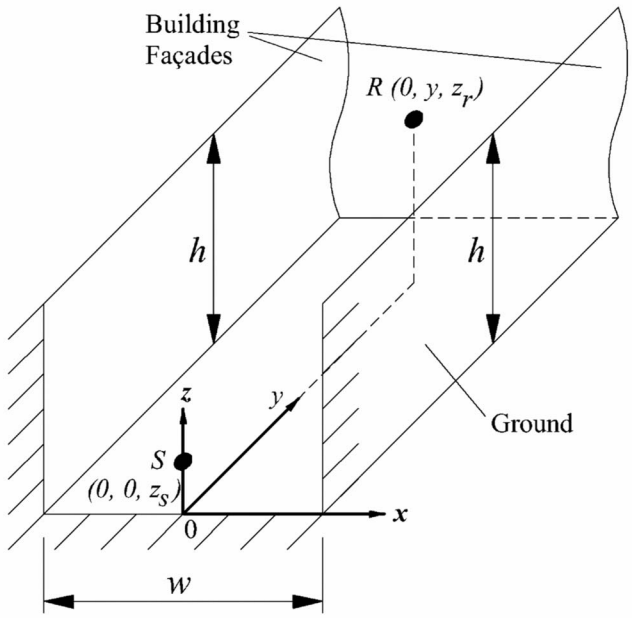


FIG. 1. Schematic showing the rectangular coordinate system and the geometrical configuration of a street canyon.

image source model where the boundaries can be taken as smooth, geometrically reflecting surfaces in the present study. It is because this model is simple yet it can be used to represent some urban situations. The closed-form analytical results can be used to offer a first order engineering approximation to supplement the numerical results obtained from other computational intensive schemes, e.g., the diffuse reflection model.[15,16]

The paper is arranged as follows. Section II presents an integral formulation to calculate the noise levels and the reverberation times in street canyons. In Sec. III, the theoretical models are validated by comparing the numerical predictions with experimental results. Section III A shows comparisons of numerical results with published experimental data.[12] In Sec. III B and III C, we present indoor and outdoor experimental results. These measured results are compared with the numerical predictions based on the integral formulations. The outcomes of the present study are summarized in Sec. IV.

II. THEORY

A. Total steady-state sound energy at a receiver

A street canyon may be considered as a category of a long enclosure: The flanking façades are treated as a pair of parallel side walls, the road pavement is taken as the floor, and the opening at the top is represented by a perfectly absorptive ceiling. Taking the boundaries as geometrically reflecting surfaces, the sound propagation along the street canyon can be calculated by means of an image source method.[16] Figure 1 displays an idealized street canyon where h is the height of the buildings measured in the z-direction. The pair of flanking façade surfaces is parallel to each other with a horizontal separation of w measured in the x-direction. The y-axis is aligned along the direction of the street canyon. The origin is located on the ground at equidistance from the parallel flanking buildings where the façade surfaces are situated at $x=\pm w/2$.[19]

In the current study, we consider a typical situation where the height of the flanking façades is greater than the width of the street canyon, i.e., $h>w$. The effect of diffraction at the top edges of the façades is ignored in the present study. It is because the total sound energy at a receiver point is dominated by the contributions from multiple reflected sound rays of the boundary surfaces. In general, the noise levels due to vehicular noise sources decrease with an increase in the receiver height because there are no reflections from the open ceiling of the street canyon. In addition, the noise levels decrease with the increase in the separation between the source and receiver along the y-direction. However, Kang[16] showed that the noise levels are relatively uniform between the buildings for a fixed horizontal separation (greater than about $1.5w$) from the source and at a constant height above the ground. Without loss of generality, we therefore assume that a point source S and a receiver R are located at the coordinates of $(0,0,z_s)$ and $(0,y,z_r)$, respectively. Hence, the sound source is separated from the receiver by a horizontal distance

$$r_I=\sqrt{y^2+(z_r-z_s)^2}. \tag{1}$$

A row of image sources, shown as the upper row in Fig. 2, is formed because of the reflections from the two vertical façade surfaces. Based on this series of image sources, the lower row of image sources can be constructed below the $z=0$ plane due to the reflection from the ground surface. All image sources can be linked to the receiver and the total sound field is computed by summing the contributions from these image sources. Making use of the image source model, the total sound energy at the receiver due to the image sources located above the ground is given by

$$\Lambda_I=\sum_{m=-\infty}^{\infty}\frac{Qe^{m\ln(1-\bar{\alpha}_v)-\alpha_a d_m}}{d_m^2}, \tag{2}$$

where Q is the source strength, $\bar{\alpha}_v$ is the mean absorption coefficient of the façade surfaces, and α_a is the air absorption factor, which can be obtained from Ref. 20 for different frequencies. The path length of the $\pm m$th image source can be determined by

$$d_m=\sqrt{(mw)^2+r_I^2}, \tag{3}$$

where r_I is determined according to Eq. (1). Here, in Eq. (2), the attenuation due to the reflection from the boundary surfaces is written in its exponential form by noting the following algebraic identity:

$$e^{m\ln(1-\bar{\alpha}_v)}=(1-\bar{\alpha}_v)^m, \tag{4}$$

where $\ln(1-\bar{\alpha}_v)<0$ because $\bar{\alpha}_v<1$ and m is the number of reflections from the façades.

Suppose that the façades are built with hard surfaces with small $\bar{\alpha}_v$. In this case, the total sound energy received at the reception point is composed of many terms due to the multiple reflections from the façade surfaces. It is possible to replace m with a continuous function in terms of x as

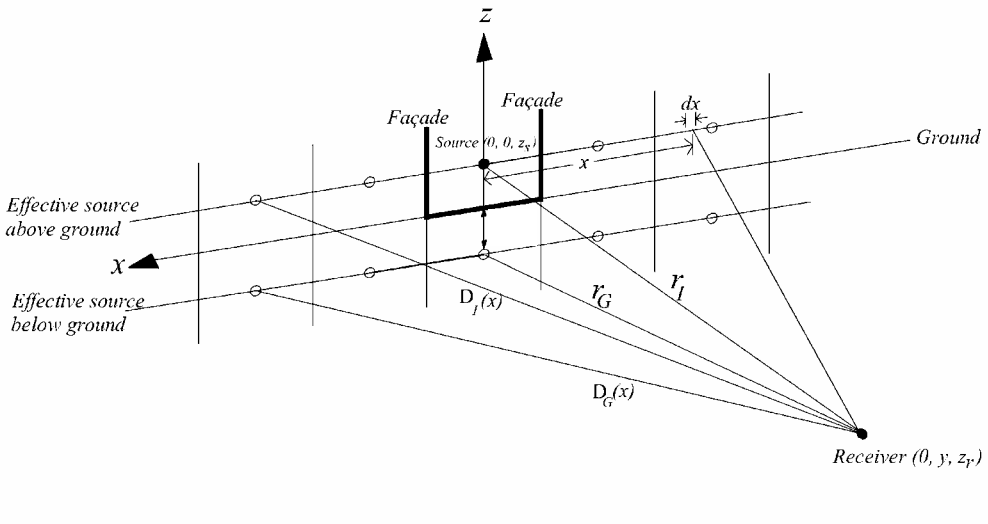


FIG. 2. Schematic to show the relative locations of the receiver, the point source (solid circle), image sources (open circle), and the corresponding line sources above and below the ground surface.

$$m = |x|/w. \tag{5}$$

A line source can now be used to replace the upper row of the discrete image sources. Figure 2 shows a schematic of the effective line source. It has an effective strength of Q/w m^{-1} because there is only one image source for the width, w, of the street canyon. The path length d_m can be replaced with

$$D_I = \sqrt{x^2 + r_I^2}. \tag{6}$$

From Eq. (2), the sound energy at the reception point due to the image sources above the ground can be written in an integral form as

$$\Lambda_I = \frac{\Lambda_0}{w}\int_0^\infty \Phi_I dx, \tag{7}$$

where Φ_I is a symmetric function of x given by

$$\Phi_I = \Phi_I(x) = \frac{2e^{\{x\ln(1-\bar{\alpha}_v)/w - \alpha_a D_I\}}}{D_I^2}, \tag{8}$$

and Λ_0 is the free-field sound energy received at a distance of 1 m from the point source. It can be determined by

$$\Lambda_0 = Qe^{-\alpha_a(1)}/1^2 \approx Q. \tag{9}$$

The integral in Eq. (7) is difficult, if not impossible, to evaluate analytically because of the presence of D_I in the exponential term of $\Phi_I(x)$. Although the integral can be evaluated by a numerical quadrature, it is more convenient to approximate the solution in an analytical form.

To this end, we approximate D_I in the exponential term of Eq. (9) by a linear function

$$D_I = \sqrt{r_I^2 + x^2} \approx r_I + K_I x, \tag{10}$$

where K_I is the slope of the approximate function. Since the solution is given in an integral form [cf. Eqs. (7) and (8)], it is convenient to approximate K_I such that

$$\int_0^{X_I} \sqrt{r_I^2 + x^2}\,dx = \int_0^{X_I} (r_I + K_I x)dx, \tag{11}$$

where the upper limit of the integrand, X_I, is set at a sufficiently large distance in order to cover the range of x that contributes to the total sound energy at the reception point. As $\Phi_I(x)$ is a monotonic decreasing function, we can approximate X_I by the following function:

$$\varepsilon = \Phi_I(X_I)/\Phi_I(0), \tag{12}$$

where ε is the ratio of the minimum sound energy and the maximum sound energy contributed due to the effective line source. The minimum sound energy is contributed from a source located at $x = X_I$ and the maximum contribution comes from the source located at $x = 0$. Preliminary numerical analyses have suggested that the predicted results are relatively the same for a wide range of ε. Hence, for simplicity, ε is chosen as 1×10^{-6} in the following numerical analyses.

It is remarkable that both integrals in Eq. (11) can be evaluated analytically to give closed-form expressions. As a result, K_I can be written in an analytical form to yield

$$K_I = \frac{2}{X_I^2}\left\{\frac{r_I^2}{2}\left[\ln\left(\frac{X_I + \sqrt{r_I^2 + X_I^2}}{r_I}\right) + \frac{X_I\sqrt{r_I^2 + X_I^2}}{r_I^2}\right] - r_I X_I\right\}. \tag{13}$$

Substituting Eqs. (8), (10), and (13) into Eq. (7), we can approximate Λ_I by

$$\Lambda_I \approx \frac{2\Lambda_0 e^{-\alpha_a r_I}}{w}\int_0^\infty \frac{e^{-\upsilon_I x}}{r_I^2 + x^2}dx, \tag{14}$$

where

$$\upsilon_I = -\ln(1 - \bar{\alpha}_v)/w + \alpha_a K_I. \tag{15}$$

By using the following identity of an indefinite integral for the exponential integral:[21]

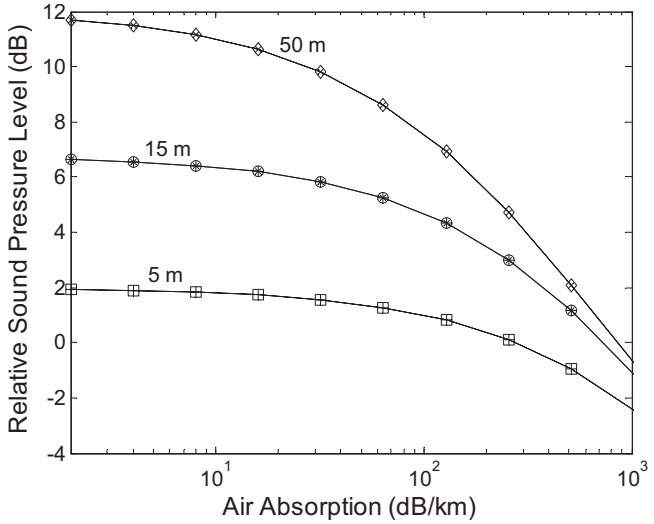


FIG. 3. The relative SPL is plotted versus the absorption factor in air. The squares, circles, and diamonds are the numerical results based on the direct numerical integration of Eq. (7) for the source/receiver separations of 5, 15, and 50 m, respectively. The plus, asterisks, and crosses are the corresponding results based on the approximate formula given in Eq. (17). The source and receiver are located at the center line of the street canyon and at the same height above the ground. The reference SPL is taken as 1 m free-field SPL; see Eq. (18).

$$\int \frac{e^x}{a^2+x^2}dx = -\frac{1}{a}\mathrm{Im}\{e^{ia}E_1(-x+ia)\} + \mathrm{const}, \tag{16}$$

we can simplify Eq. (14) to yield

$$\Lambda_I \approx \frac{-2\Lambda_0 e^{-\alpha_a r_I}}{w r_I}\mathrm{Im}[e^{i\upsilon_I r_I}E_1(i\upsilon_I r_I)], \tag{17}$$

where E_1 is the exponential integral with complex arguments.

We demonstrate the accuracy of Eq. (17) by comparing the numerical results obtained by direct numerical integration of Eq. (7). Figure 3 shows comparisons of these two numerical results in which the relative sound pressure level (SPL) is plotted against the air absorption factor α_a. In the plots, the relative SPL is defined as

$$\mathrm{SPL}_r = 10\log(\Lambda_I/\Lambda_0), \tag{18}$$

where Λ_0 is the free-field sound energy at 1 m from the source. In Fig. 3, a street canyon of 3 m wide and 100 m high is used in the numerical calculations. We assume that the building façades are perfectly reflecting surfaces, i.e., $\bar{\alpha}_v=0$. Numerical results with the horizontal separation between the source and receiver of 5, 15, and 50 m, respectively, are shown. Both source and receiver are located at equidistance from façades and at the same height above the ground. It follows from Fig. 3 that the numerical results according to both schemes agree very well with each other. These comparisons have demonstrated the accuracy of the approximate model, which is sufficient to estimate the effect of the atmospheric absorption in a street canyon of typical source/receiver geometrical configurations.

Similarly, we can derive an analogous expression for the sound energy due to the image sources located below the ground as follows:

$$\Lambda_G = \frac{\Lambda_0 e^{-\alpha_a r_G}}{w}\int_0^\infty \Phi_G dx, \tag{19}$$

where the subscript G denotes the corresponding parameters for the ground-reflected waves. The symmetric function $\Phi_G(x)$ is given by

$$\Phi_G = \Phi_G(x) = \frac{2(1-\alpha_G)e^{\{x\ln(1-\bar{\alpha}_v)/w-\alpha_a D_G\}}}{D_G^2}, \tag{20}$$

where α_G is the absorption coefficient of the ground, D_G is the total distance between the receiver and the image sources below the ground

$$D_G = \sqrt{x^2+r_G^2}, \tag{21}$$

and r_G is the corresponding horizontal separation

$$r_G = \sqrt{y^2+(z_r+z_s)^2}. \tag{22}$$

Using an analogous approach, we can derive an approximate expression for the sound energy at the reception point due to the image sources below the ground as follows:

$$\Lambda_G = \frac{-2\Lambda_0(1-\alpha_G)e^{-\alpha_a r_G}}{w r_G}\mathrm{Im}[e^{i\upsilon_G r_G}E_1(i\upsilon_G r_G)], \tag{23}$$

where the parameters υ_G and K_G are given by

$$\upsilon_G = -\ln(1-\overline{\alpha_v})/w + \alpha_a K_G \tag{24a}$$

and

$$K_G = \frac{2}{X_G^2}\left\{\frac{r_G^2}{2}\left[\ln\left(\frac{X_G+\sqrt{r_G^2+X_G^2}}{r_G}\right)+\frac{X_I\sqrt{r_G^2+X_G^2}}{r_G^2}\right] - r_G X_G\right\}. \tag{24b}$$

By means of the same approach, it is possible to determine X_G by solving an analogous equation [cf. Eq. (11)] numerically. However, it is remarkable that the same term (i.e., $X_G=X_I$) may also be used in Eq. (24b) for calculating K_G. A numerical analysis has suggested that a small variation in X_G does not cause a significant change in the numerical values of Λ_G. The details of these numerical analyses are not shown here for brevity.

The total sound energy can now be found by summing the contributions from all image sources (above and below the ground) to give

$$\Lambda_T = \Lambda_I + \Lambda_G, \tag{25}$$

where Λ_I and Λ_G are determined by Eqs. (7) and (19), respectively. It is possible to give a closed-form solution for Λ_T by substituting Eqs. (17) and (23) into Eq. (25) to yield

$$\Lambda_T = \frac{-2\Lambda_0 e^{-\alpha_a r_I}}{w r_I}\mathrm{Im}[e^{i\upsilon_I r_I}E_1(i\upsilon_I r_I) + A_G(1-\alpha_G)e^{i\upsilon_G r_G}E_1(i\upsilon_G r_G)], \tag{26}$$

where A_G is a correction factor for the change in distances because the image sources are located below the ground. It is given by

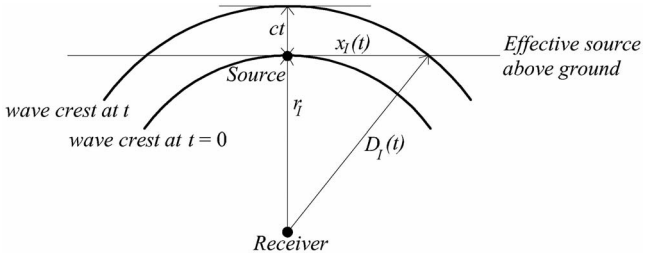


FIG. 4. Schematic to show the wave crests centering at the receiver at time $t=0$ and at time t.

$$A_G = (r_I/r_G)e^{-\alpha_a(r_G-r_I)}. \tag{27}$$

If the source is located near the ground and at a large horizontal separation from the receiver (i.e., $r_I \approx r_G$), A_G is approximately equal to unity.

B. Reverberation time in street canyons

To assess the effect of multiple reflections, it is useful to determine the reverberation times in a street canyon. The problem may be treated as the determination of the transient sound field when the source is either switched on or turned off. According to the principle of reciprocity, we may consider the receiver as the center of the wave fronts: the crests of spherical waves radiating from the receiver, which arrive at different image sources at different times. For convenience, the initial time $t=0$ is set at the moment when the direct sound wave reaches the receiver R. Suppose that the source is turned on at time $t=-r_I/c$. Then the position of a wave crest at time $t>0$ is located at a horizontal distance $x_I(t)$ from the source S. By a simple geometrical consideration, the path length D_I from the receiver to an image source above the ground can be determined by

$$D_I(t) = \sqrt{x_I(t)^2 + r_I^2} = r_I + ct, \tag{28}$$

where r_I is given by Eq. (1); see Fig. 4 for a schematic of the wave crests at time $t=0$ and at time t.

According to Eq. (14), the transient sound energy due to the image sources above the ground can then be written as

$$\Gamma_I(t) = \frac{2\Lambda_0 e^{-\alpha_a r_I}}{w}\int_0^{x_I(t)} \frac{e^{-\upsilon_I x}}{r_I^2 + x^2}dx, \quad t>0, \tag{29}$$

since only those sources located in the region $-x_I(t) \le x \le x_I(t)$ can contribute to the sound field at the reception point. Equation (29) can be reduced to Eq. (14) when $t\to\infty$, i.e., the total transient sound energy $\Gamma_I(\infty)$ becomes the steady-state sound energy at the reception point.

We can also consider a complimentary situation when the source is originally turned on but it is deactivated at $t=-r_I/c$ where c is the sound speed in air. In this case, the transient sound energy at time t is simply given by

$$\Gamma_I^*(t) = \frac{2\Lambda_0 e^{-\alpha_a r_I}}{w}\int_{x_I(t)}^{\infty} \frac{e^{-\upsilon_I x}}{r_I^2 + x^2}dx, \quad t>0. \tag{30}$$

The integrals given in Eqs. (30) and (29) can be evaluated, respectively, to yield

$$\Gamma_I^*(t) = -\frac{2\Lambda_0 e^{-\alpha_a r_I}}{wr_I}\mathrm{Im}\{e^{i\upsilon_I r_I}E_1[\upsilon_I(x_I + ir_I)]\} \tag{31}$$

and

$$\Gamma_I(t) = \Lambda_I - \Gamma_I^*(t), \tag{32}$$

where Λ_I is determined from Eq. (17).

Similarly, the transient sound energy due to the image source located below the ground can be determined when the source is turned off at $t=-r_I/c$. However, the sound energy starts to decay at a later time at $t=(r_G/r_I)/c$ because r_G/r_I. The transient sound energy is given by

$$\Gamma_G^*(t) = -\frac{2\Lambda_0 e^{-\alpha_a r_G}}{wr_G}\mathrm{Im}\{e^{i\upsilon_G r_G}E_1[\upsilon_G(x_G + ir_G)]\}, \quad t \ge (r_G - r_I)/c, \tag{33}$$

where x_G is determined by the following equation:

$$D_G(t) = \sqrt{x_G(t)^2 + r_G^2} = r_G + ct. \tag{34}$$

On the other hand, if the source is activated at $t=0$, then the transient sound energy is given by

$$\Gamma_G(t) = \Lambda_G - \Gamma_G^*(t), \quad t \ge (r_G - r_I)/c. \tag{35}$$

The effect of multiple reflections on street canyons can be assessed by considering the reverberation times, T_{60} and T_{30}, where T_{60} is the time for the noise level to reduce by 60 dB below the initial level and T_{30} is the decay time for the noise level to drop from -5 to -35 dB. A decay curve in the noise level is normally needed to determine T_{60} and T_{30}. Therefore, we focus on the transient sound energy when the source is turned off in favor of the transient sound energy when the source is activated. The total transient sound energy can be obtained by summing the contributions from the image sources located above and below the ground. From Eqs. (31) and (33), we can obtain

$$\begin{aligned}\Gamma_T^*(t) &= \Gamma_I^*(t) + \Gamma_G^*(t) \\ &= -\frac{2\Lambda_0 e^{-\alpha_a r_I}}{wr_I}\mathrm{Im}\{e^{i\upsilon_I r_I}E_1[\upsilon_I(x_I + ir_I)] \\ &\quad + A_G(1-\alpha_G)e^{i\upsilon_G r_G}E_1[\upsilon_G(x_G + ir_G)]\}.\end{aligned} \tag{36}$$

Since the current formulation is based on the image source method, it is important to demonstrate the validity of the model by comparing the current numerical results with those predicted by the image source model. It is sufficient to show plots of the decay curves with the source turned off at $t=r_I/c$. Figure 5 displays these comparisons at different mean absorption coefficients of all boundary surfaces of 0.15, 0.2, 0.25, and 0.3, respectively. The width of 10 m and the source/receiver separation of 10 m are used in the calculations. In the graphs, the relative SPLs are defined as

$$L(t) = 10\log[\Gamma_T^*(t)/\Gamma_T^*(0)], \tag{37}$$

where the reference level $\Gamma_T^*(0)$ is taken as the initial noise level. The source and receiver are placed, respectively, at 0.5 and 5.0 m above the ground and they are located at equidistance from the façade surfaces. We can see from Fig. 5 that the predicted decay curves according to the image source

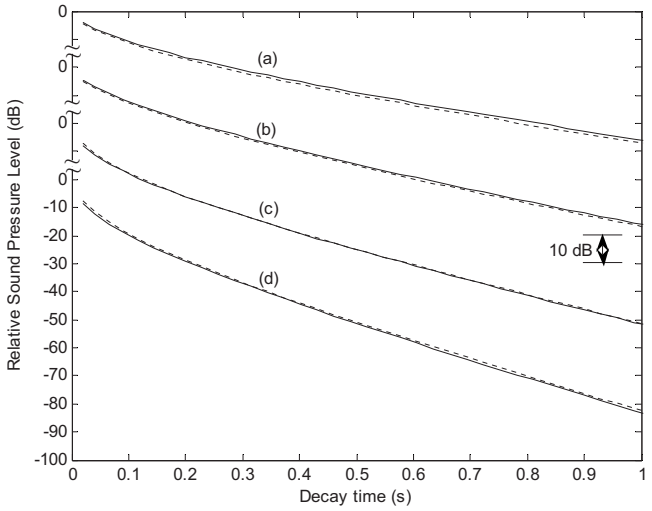


FIG. 5. The relative SPL is plotted versus time in a street canyon with the width of 10 m, and the source/receiver separation of 10 m. The source and receiver are placed, respectively, at 0.5 m and receiver at 5 m above the ground and they are located at equidistance from the façade surfaces. The mean absorption coefficients of all boundary surfaces are the same at (a) 0.15, (b) 0.2, (c) 0.25, and (d) 0.3. The solid lines (——) are the predictions by the integral formulation and the dotted lines (-----) are the predictions by the image source method. The reference SPL is taken at the initial time when the source is turned off; see Eq. (37).

method and the integral formulation agree tolerably well with each other in all cases.

Using Eq. (36), the decay curve can be produced by a reverse-time integration of the transient response. This process is known as the Schroeder approach.[22,23] The reverberation time T_{60} is then determined by using the rate of decay given by the linear regression with the noise level reducing by 60 dB from the initial level. In other words, T_{60} can be determined for the time when the ratio of $\Gamma_T^*(t)/\Gamma_T^*(0)$ is reduced to 10^{-6}. If the reverberation time T_{30} is required, Eq. (37) will be used to obtain the decay curve. A linear regression analysis is conducted to determine the rate of decay for the noise level varying between 5 and 35 dB below $\Gamma_T^*(0)$.

III. COMPARISON WITH EXPERIMENTAL MEASUREMENTS

A. Full scale field measurements in a town street

In the 1970s, Steenackers *et al.*[12] carried out a series of measurements for the reverberation time of typical town streets with their widths varying between 9 and 60 m. An alarm gun was used as an impulsive noise source where the sound decay curves were measured. The reverberation time was then used to estimate the sound absorption coefficient of the town street. Their experimental data[12] will be used to validate the integral expression given in Eq. (36). We also compare the numerical results based on the image source model with geometrically reflected boundary surfaces.[16]

Figure 6 displays the measured time histories of the relative SPLs adapted from the published data.[12] They were the sound level decay curves in the town streets with the respective widths of 12, 18, and 40 m. The horizontal distances between the source and receiver for all measurements, which were estimated from the experimental data, were 10 m. Since the exact locations of the source and receiver are not available in Ref. 12, we assume that they were placed at 1.2 m above the ground and at equidistance from the flanking buildings.

To confirm the validity of this assumption, a prior numerical analysis is conducted to examine the effect of the source location on the SPLs at different receiver locations along the street canyon. It is found that the sound fields are generally uniform within the same cross sections of the street canyon, especially, when the source and receiver are close to the ground and their horizontal separation is sufficiently

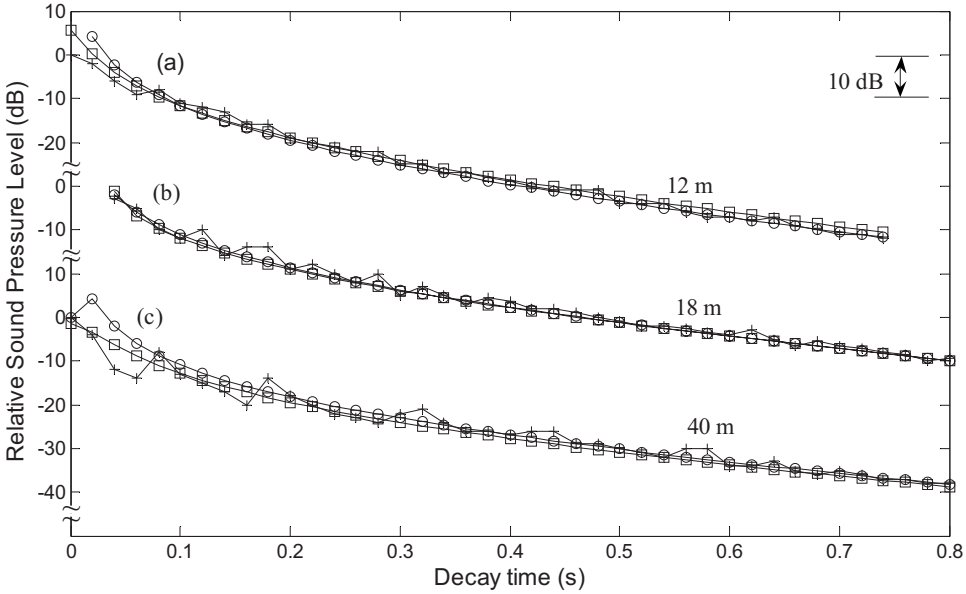


FIG. 6. The relative SPL is plotted versus time in town streets with the source/receiver separation of 10 m and the source and receiver were located at 1.2 m above the ground. The widths of the town street are (a) 12, (b) 18, and (c) 40 m. The lines with plus signs (+) are measurement results taken from Ref. 12. The lines with squares (□) are predictions by the integral formulation. The lines with circles (○) are predictions by the image source method. The reference SPL is taken at the initial time when the source is turned off; see Eq. (37).

large. These numerical simulations are not shown here but this finding is in agreement with the conclusion suggested by Kang.[16] With these prior numerical simulations, we expect that their assumption on the source/receiver locations will be sufficient to give accurate predictions of the sound fields in the street canyons.

In addition to the published data, Fig. 6 also presents the numerical predictions according to the image source model[16] and the current formulation. In his numerical analyses, Kang[16] also used the incoherent model but he computed the overall SPL by summing the contributions from all image sources, which have been "switched" on. The number of image sources increases discretely to its steady-state level after the first arrival of the direct wave. The number of "switched-on" image sources reaches its maximum level after a short duration. Hence, the image source model predicts a step change in the sound energy level shortly after the first arrival time but time-varying sound energy will gradually become the steady-state level as the time increases.

In the numerical calculations of all town streets, we assume that the air absorption factor is zero because no pertinent information is provided for the atmospheric conditions during the measurement periods. In Fig. 6(a), we present the predicted and measured results for the town street of 12 m wide. According to Steenackers *et al.*,[12] the ground was assumed to be a perfectly reflecting surface and the average mean absorption coefficient of the building façades was 0.15. For the given source and receiver locations, Eq. (37) is used to evaluate the time history of the relative SPLs. The numerical predictions according to the image source model and the integral formulation agree to within 1.1 dB in the relative SPL. Furthermore, both predicted results show good agreements with the experimental data.

Figure 6(b) presents a set of data taken from Fig. 1 of Ref. 12 where the width of the town street was 18 m. The estimated sound absorption coefficient, which was taken from Table I of Ref. 12, was 0.17. Again, we use Eq. (37) to calculate the time histories of the relative SPLs. The average discrepancy between the measured data and the image source model is 1.21 dB while it is 1.15 dB for the integral formulation.

The compared results for the last set of data are displayed in Fig. 6(c). Based on the information obtained from Steenackers *et al.*,[12] we estimated that the mean absorption coefficient of the façade was 0.3. Figure 6(c) shows tolerably good agreements between the measured data and the numerical results according to both prediction methods. The average discrepancies between the image source method and measured results are 1.54 and 1.35 dB for that of the integral formulation.

By comparing with published experimental data, it is enlightening to confirm the validity of the analytical formulation for predicting the relative SPL in a street canyon. The proposed model provides a simple closed-form solution, which compares well with the standard image source model and with the published experimental results.

FIG. 7. (Color online) Photograph showing the setup of the indoor experiment conducted in an anechoic chamber.

B. Indoor model experiments

A model street canyon was built and placed in an anechoic chamber of internal dimensions of $6\times6\times4$ m^3 (high). Hard plywood boards of 20 mm thickness were used to construct the model street canyon of 0.8 m wide, 5 m long, and 2.6 m high. The hard plywood boards were varnished to prevent leakage of sound. To simulate the façade surfaces with finite impedance, the two vertical walls were covered with acrylic light diffusers of 5 mm thickness; see Fig. 7. The light diffusers have rectangular grids with dimensions of 15×15 mm^2 and open volumes. The light diffusers may give diffused sound fields but, nevertheless, Daigle *et al.*[24] demonstrated experimentally that a hard ground covered with light diffusers can be used to simulate an indoor impedance plane. In light of the suggestion of Daigle *et al.*,[24] we conducted a set of short-range measurements for the propagation of sound above the hard plywood board covered with and without light diffusers. In these measurements, the source and receiver were placed at a horizontal separation of 1 m and at heights of 0.01 m above the model ground.

Figure 8(a) shows a typical measured spectrum to demonstrate the interference effects due to the direct and reflected waves for sound propagation above the model impedance ground. The excess attenuation, which is defined as the ratio of the total field above the impedance surface to the free-field measurement at 1 m in the anechoic chamber, is plotted against the source frequency in Fig. 8(a). We also conducted similar measurements for the propagation of sound over the hard plywood board. Its measured excess

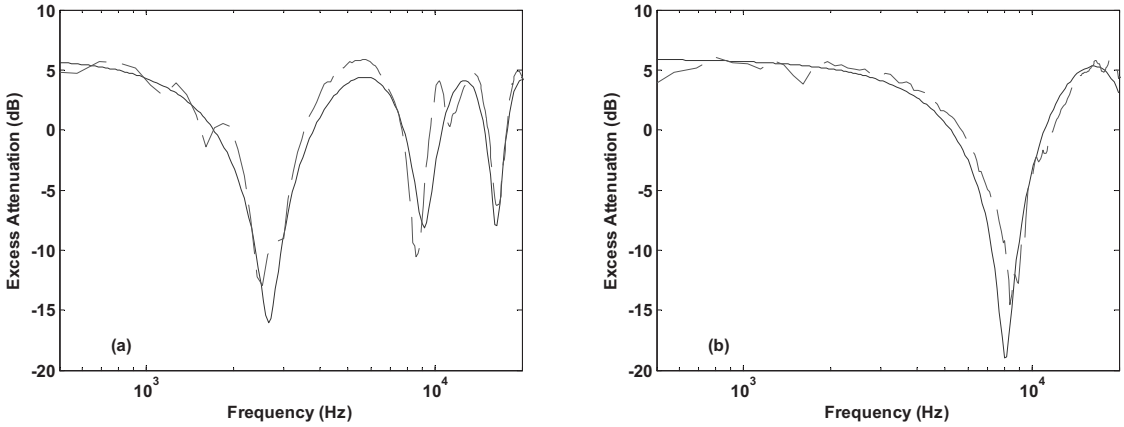


FIG. 8. The acoustic characterization of (a) the vertical walls (indoor façades) and (b) the hard floor (indoor ground). The two-parameter model was used to characterize the surface impedances. These parametric values are listed in Table I for information. The solid lines are (——) are theoretical predictions and the dotted lines (----) are experimental data.

attenuation spectrum is shown in Fig. 8(b). From both graphs in Fig. 8, there are clear interference dips due to the interference of the direct and reflected waves although the primary dip occurs at higher frequency for the hard ground. It is because the impedance of a hard surface is usually higher than that of the model plane with finite impedance. From these two sets of short-range measurements in the anechoic chamber, we can confirm the suitability of treating the plywood board covered with light diffusers as a model impedance plane.

We also use the measured excess attenuation spectra in Fig. 8 to determine the acoustic impedances of the plywood board with (impedance façade surfaces) and without (hard ground surfaces) the cover of light diffusers. Attenborough's two-parameter model[25] was used to characterize the surface impedances, Z, of the vertical wall and the ground in the present study. The surface impedance is calculated by

$$Z = 0.538\sqrt{\sigma_e/f} + i[0.538\sqrt{\sigma_e/f} + 19.74\alpha_e/f], \tag{38}$$

where f is the source frequency, σ_e is the effective flow resistivity, and α_e is the effective rate of change in porosity with depth. The parametric values of σ_e and α_e are deduced from the short-range propagation measurement over the surface.

The best-fit parametric values for σ_e and α_e were 80 kPa s m^{-2} and 450 m^{-1} for the vertical walls (plywood boards covered with light diffusers) and 80 000 kPa s m^{-2} and 50 m^{-1} for the ground made of the plywood board. These parametric values for different boundary surfaces used in the numerical simulations are summarized in Table I for the ease of reference. Figure 8 also shows typical predicted excess attenuation spectra for the propagation of sound above the vertical walls and the ground where their impedances are calculated by using Eq. (38) with the parametric values given in Table I.

For the image source model and the current integral formulation, the one-third octave band absorption coefficients (α) of the boundary surfaces are calculated from the measured impedance by[22]

$$\alpha = \left| \frac{Z-1}{Z+1} \right|^2. \tag{39}$$

Using the Attenborough two-parameter impedance model,[25] the absorption coefficients of the ground and vertical walls in one-third octave bands are calculated and listed in Table II.

In indoor measurements, a Renkus-Heinz PN 61 self-powered loudspeaker was used as the source and a "B&K" 4189 pre-polarized free-field 1/2 in. condenser microphone was used as the receiver. A PC-based maximum length sequence system analyzer (Ref. 26) was used as both signal generator for the source and data analyzer of the measurements. The source strength was characterized by conducting prior measurements to measure its SPL at 1 m free field in the anechoic chamber. In the present indoor experiments, we presented the results for steady-state SPLs at various horizontal distances at different frequency bands. The source was located at 0.545 m away from the left vertical wall and at heights varying between 0.10 and 2.1 m. For the receiver, it was also situated at 0.545 m from the left vertical wall and 0.2 m above the ground.

Figure 9 shows the typical spectra of the steady-state SPLs where the receiver was located at 2 and 4 m from the source. The source and received were, respectively, placed at 0.6 and 0.2 m above the ground and at the same distance of 0.545 m from the left vertical wall. In the graphs, we compare the measured spectra with the numerical predictions according to the integral formulation. The measured and nu-

TABLE I. The best-fit parametric values of the effective flow resistivity (σ_e) and the effective rate of change in porosity with depth (α_e) for predicting the acoustic impedance of the boundary surfaces used in the indoor and outdoor experiments.

	σ_e (kPa s m^{-2})	α_e (m^{-1})
Indoor façade	80	450
Indoor ground	80 000	50
Outdoor façade	50 000	500
Outdoor ground	200 000	500

TABLE II. The estimated mean absorption coefficients of the boundary surfaces given in dB and the absorption coefficient of air at 30 °C and 80% relative humidity given in dB/km.

Frequency (Hz)	400	500	630	800	1000	1250	1600	2000	2500	3150	4000	5000	6300	8000
Absorption coefficient (indoor façade)	0.07	0.09	0.11	0.14	0.18	0.21	0.25	0.29	0.32	0.34	0.36	0.37	0.36	0.35
Absorption coefficient (indoor ground)	0.026	0.028	0.032	0.036	0.04	0.045	0.05	0.056	0.062	0.07	0.079	0.088	0.098	0.11
Absorption coefficient (outdoor façade)	0.028	0.03	0.036	0.041	0.046	0.052	0.059	0.067	0.075	0.084	0.094	0.11	0.12	0.13
Absorption coefficient (outdoor ground)	0.011	0.013	0.014	0.016	0.018	0.020	0.023	0.026	0.028	0.032	0.036	0.04	0.045	0.05
Air absorption (30 °C, 80% R.H.) (dB/km)	1.98	2.79	4.14	5.66	7.41	9.28	11.2	13.3	15.7	18.8	23.1	29.7	39.8	55.7

merical predictions agree reasonably well with each other except for one or two individual one-third octave bands, e.g., a discrepancy of about 5 dB was found for 3150 Hz in the 2 m separation and about 7 dB for 5 kHz in the 4 m separation. It is also observed that the integral formulation can predict the general trend of the measured spectral SPLs. The average discrepancies in the one-third octave bands were found to be 1.9 and 2.7 dB, respectively, for the 2 and 4 m separations.

The variation of the steady-state SPL with the horizontal separation is shown in Fig. 10 where four sets of typical results for the frequencies of 500 Hz, 1.25 kHz, 2.5 kHz, and 5 kHz are presented. The relative source/receiver positions with the ground and with the vertical walls are the same as in Fig. 9. These plots serve to highlight the capability of the integral formulation in predicting the general trend of measured data for the indoor model experiments.

Before we end this section, they wish to point out that there were other studies[26,27] examining an improved technique for simultaneously selecting both an optimal scale factor and optimal model materials for use in indoor scale model experiments. However, there is no attempt to select the most appropriate materials for modeling the façade surfaces in the present study. The use of a model impedance plane is sufficient to validate the integral formulation by comparing its predicted SPLs with the precise indoor measurements.[28]

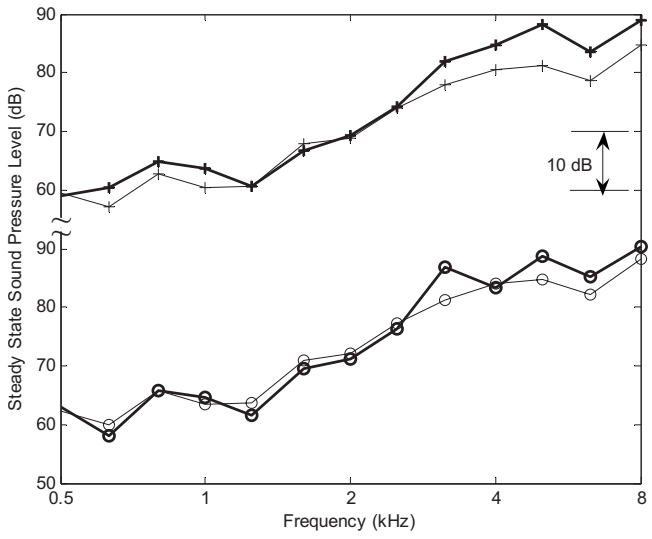


FIG. 9. The SPL spectra for the indoor model experiments. The source was located at 0.6 m above the ground and at 0.565 m from one side of the vertical wall. The receiver was placed at 0.2 m above the ground and 0.215 from the same wall. The thick solid lines are experimental data and the thin solid lines are predictions by the integral formulations. The lines with open circles and plus signs are for the horizontal separations of 2 and 4 m, respectively.

C. Outdoor field measurements in an alley street

To confirm the versatility of the integral formulation, we conducted a further set of full scale field measurements in an alley street. It was 3.13 m wide, 20 m long, and the height of the two parallel was 54 m. The alley street had a concrete ground and both vertical walls were covered with hard mosaic tiles. We used the same set of instrument for all indoor and outdoor measurements. Figure 11 shows the experimental setup in the alley street. In the experiments, the source was located at the center of the alley street and 0.1 m above the ground. The receiver was placed at a height of 1.6 m above the ground and at 0.69 m from one of the vertical walls. The distance between source and receiver was adjusted between 1 and 20 m.

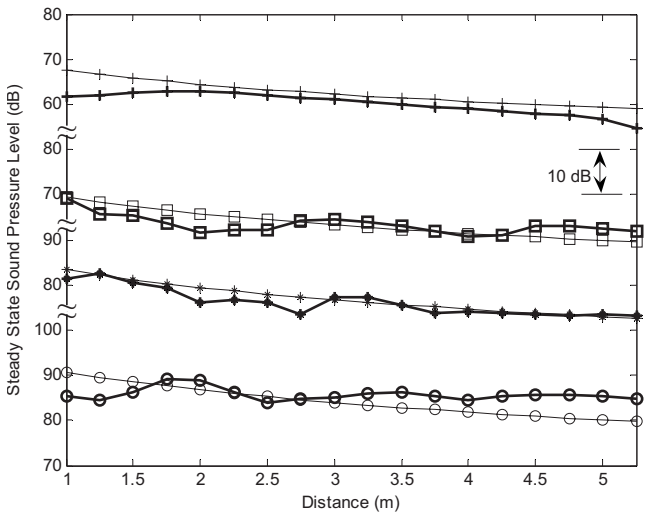


FIG. 10. The SPL is plotted versus the horizontal distance. The source and receiver locations are the same as Fig. 9. The thick solid lines are for experimental data and thin solid lines are for predictions by the integral formulation. (Plus signs: 500 Hz, squares: 1.25 kHz, asterisks: 2.5 kHz, and open circles: 5 kHz.)

FIG. 11. (Color online) Photograph showing the experiment setup of the field measurements in an alley street.

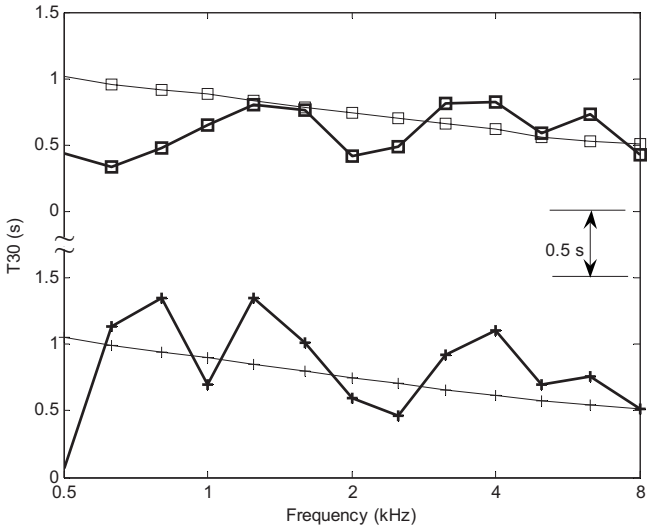


FIG. 12. The reverberation time T_{30} is plotted versus the source frequency where the source is placed at 0.1 m above the ground and at 1.56 m from one side of the vertical wall. The receiver was located at 1.6 m above the ground and 0.69 from the same vertical wall. The thick solid lines are for experimental data and thin solid lines are for predictions by the integral formulation. (Plus signs: source/receiver separation of 4 m and squares: source/receiver separation of 12 m.)

Prior *in-situ* measurements were conducted to determine the sound absorption coefficient of the vertical walls and the ground. The source and receiver were placed at 0.01 m above the respective surfaces and their horizontal separation was set at 1 m in these measurements. Again, the Attenborough two-parameter impedance model[25] was used to characterize the respective impedances of the surfaces and their parametric values are listed in Table I. All these boundary surfaces were made of acoustically hard construction materials, which led to fairly high parametric values for σ_e according to the short-range characterization measurements.

The sound absorption coefficients of the concrete ground and the vertical walls covered with mosaic tiles can then be calculated according to Eq. (39). These absorption coefficients, which are listed in Table II, will then be used in subsequent calculations of the reverberation times and SPLs in the alley street.

During the measurement, the outdoor temperature was about 30 °C and the relative humidity was 80%, the air absorption factors which are used in the following predictions and are obtained from Ref. 20. The air absorption coefficients at different frequency bands are listed in Table II. Due to the relatively high background noise levels in the field measurements, the reverberation time T_{30} was measured in favor of T_{60}. The SPL of the loudspeaker was measured at 1 m free field in the anechoic chamber before the field measurements. It was used to compare with the SPLs measured in the alley street.

Figure 12 shows the experimental data of T_{30} in one-third octave bands for the source and receiver located along the alley street. We compare the measured data with the numerical predictions of reverberation times for horizontal separations of 4 and 12 m. The numerical results according to the integral formulations are also shown in Fig. 12. We remark that the numerical predictions can only give a reasonable estimation of the general trend of T_{30} as the source frequency increases. Comparison of results for other source/receiver separations was generally similar to those shown in Fig. 12 but they are not shown here for succinctness. In this set of data, the source and receiver are placed at respective heights of 0.1 and 1.6 m above the ground. They are located at respective distances of 1.56 and 0.69 m from one of the vertical walls.

Figure 13 displays the variations of reverberation times with the horizontal separation. The source/receiver positions relative to the vertical walls and the ground are the same as above. Numerical results and measured data are presented for the source frequencies of 800 Hz, 1.6 kHz, 2 kHz, and 4 kHz. Similar results are also obtained for other frequencies but only four frequencies are selected for presentation. As shown in the plots, the predicted T_{30} varies marginally along the horizontal range (up to 15 m) although the experimental

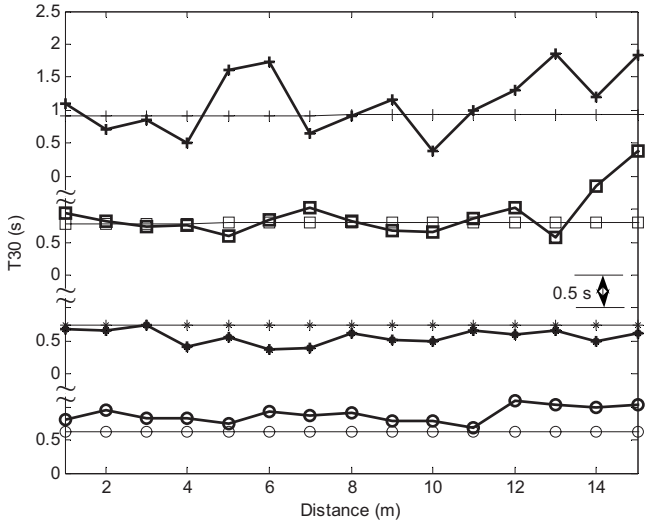


FIG. 13. The reverberation time T_{30} is plotted versus the distance. The source and receiver have the same relative positions from the vertical walls and the ground as described in Fig. 12. The thick solid lines are for experimental data and thin solid lines are for predictions by the integral formulation. (Plus signs: 800 Hz, squares: 1.6 kHz, asterisks: 2 kHz, and open circles: 4 kHz.)

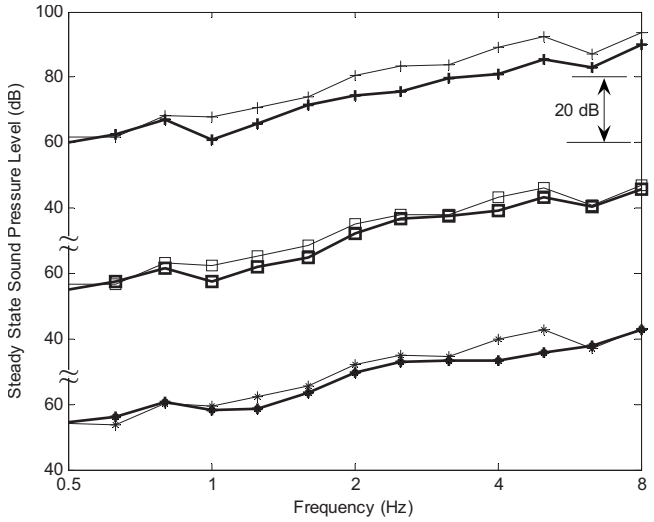


FIG. 14. The steady-state SPL is plotted versus the source frequency with the same source/receiver geometry as Fig. 12. The thick solid lines are for experimental data and thin solid lines are for predictions by the integral formulation. The respective source/receiver separations are 4 (plus signs), 12 (squares), and 20 m (asterisks).

data show a fluctuation in T_{30}. In general, the integral formulation can predict the average level of measured T_{30} over the range of interest.

Next, the same source/receiver geometries as above are used for the next two sets of the experimental measurements. In the first set of data, the predicted and measured spectra for the steady-state SPLs are compared in Fig. 14 for the respective horizontal separations of 4, 12, and 20 m. Apparently, the numerical predictions according to the integral formulation give reasonably good agreements with the general trends of the measured frequency spectra in all cases. In the second set of data, the variations of steady-state SPLs with the horizontal range were shown in Fig. 15 for the respective frequencies of 500 Hz, 800 Hz, 2 kHz, and 4 kHz. Again, the predicted results show tolerably good agreements with the general trend for the reduction in the SPLs when the separation between the source and receiver increases.

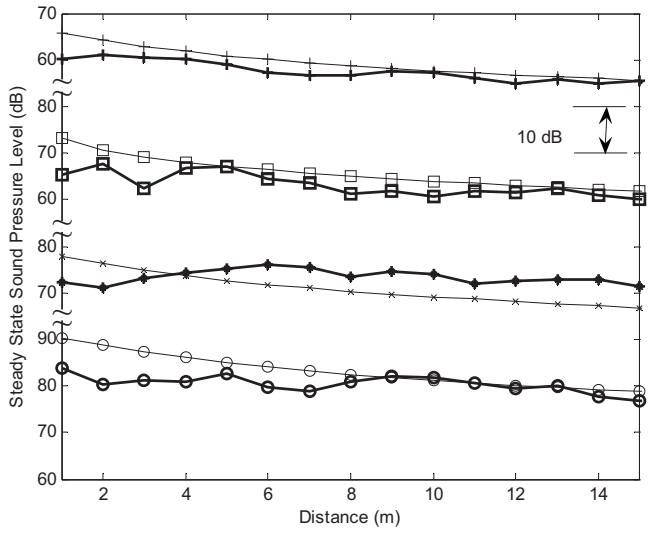


FIG. 15. The steady-state SPL is plotted versus the distance. The source/receiver geometry is same as Fig. 13. The thick solid lines are for experimental data and thin solid lines are for predictions by the integral formulation. (Plus signs: 500 Hz, squares: 800 Hz, asterisks: 2 kHz, and open circles: 4 kHz.)

IV. CONCLUSIONS

By modeling a street canyon as a long enclosure with an open top, it is possible to use an energy approach to sum the contribution from all image sources incoherently. In the present study, an integral formulation has been developed to estimate the noise levels in a street canyon by replacing the discrete image sources with an effective line source. The integral formulation is used to predict the decay curve of sound energy where the reverberation time in the street canyon can be estimated. Based on the reverberation time, the steady-state SPLs in the street canyon can be calculated. It has been demonstrated that the predictions according to the integral formulation agreed reasonably well with published data conducted in a town street. The numerical results according to the integral formulation also show good agreements with indoor and outdoor experimental data obtained in the present study. The integral formulation can be used to provide an efficient model for predicting noise levels and the reverberation effect in a street canyon.

ACKNOWLEDGMENTS

We are grateful to S. T. So, William Fung, and T. L. Yip for their help in conducting the outdoor and indoor experiments reported in this study. We are grateful to the three anonymous reviewers for their helpful comments.

[1]C. H. Chew, "Prediction of traffic noise from expressways—Part 1: Buildings flanking one side of expressways," Appl. Acoust. **28**, 203–212 (1989).
[2]C. H. Chew, "Prediction of traffic noise from expressway—Part II: Buildings flanking both sides of expressways," Appl. Acoust. **32**, 61–72 (1991).
[3]K. K. Chin, K. Y. Ng, S. K. Soon, and H. T. Chui, "A study of urban traffic noise," Proceedings Western Pacific Regional Acoustics Conference **2**, 142–145 (1997).
[4]K. M. Li, "Managing environment noise in Hong Kong," J. Acoust. Soc. Am. **115**, 2593 (2002).
[5]J. E. Volkmann and M. L. Graham, "A survey on air raid alarm signals," J. Acoust. Soc. Am. **14**, 1–9 (1942).
[6]L. M. Ball, "Air raid siren field tests," J. Acoust. Soc. Am. **14**, 10–13 (1942).
[7]F. W. Wiener, C. I. Malme, and C. C. Gogos, "Sound propagation in urban areas," J. Acoust. Soc. Am. **37**, 738–747 (1965).
[8]D. Aylor, J. Y. Parlange, and C. Chapman, "Reverberation in a city street," J. Acoust. Soc. Am. **54**, 1754–1757 (1973).
[9]K. W. Yeow, "External reverberation times observed in built-up area," J. Sound Vib. **48**, 438–440 (1976).
[10]K. W. Yeow, "Decay of sound levels with distance from a steady source observed in a built-up area," J. Sound Vib. **52**, 151–154 (1977).
[11]N. W. M. Ko and C. P. Tang, "Reverberation time in a high-rise city," J. Sound Vib. **56**, 459–461 (1978).
[12]P. Steenackers, H. Myncke, and A. Cops, "Reverberation in town streets," Acustica **40**, 115–119 (1978).
[13]R. H. Lyon, "Roles of multiple reflections and reverberation in urban noise propagation," J. Acoust. Soc. Am. **55**, 493–503 (1974).
[14]H. G. Davies, "Multiple-reflection diffuse-scattering model for noise propagation in streets," J. Acoust. Soc. Am. **64**, 517–521 (1978).
[15]J. Picaut, L. Simon, and J. Hardy, "Sound field modeling in streets with a diffusion equation," J. Acoust. Soc. Am. **106**, 2638–2645 (1999).
[16]J. Kang, "Sound propagation in street canyons: Comparison between diffusely and geometrically reflecting boundaries," J. Acoust. Soc. Am. **107**, 1394–1404 (2000).

[17]J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small room acoustics," J. Acoust. Soc. Am. **65**, 943–950 (1979).
[18]K. K. Lu and K. M. Li, "The propagation of sound in narrow street canyons," J. Acoust. Soc. Am. **112**, 537–550 (2002).
[19]J. Kang, "Numerical modeling of the sound fields in urban streets with diffusely reflecting boundaries," J. Sound Vib. **258**, 793–813 (2002).
[20]"Method for calculation of the absorption of sound by the atmosphere," American National Standards Institute S1.26 (1995).
[21]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions, With Formulas, Graphs and Mathematical Tables* (Wiley, New York, 1972), Chap. 5.
[22]K. M. Li and P. M. Lam, "Prediction of reverberation time and speech transmission index in long enclosures," J. Acoust. Soc. Am. **117**, 3716–3726 (2005).
[23]M. R. Schroeder, "New method of measuring reverberation time," J. Acoust. Soc. Am. **37**, 409–412 (1965).
[24]G. A. Daigle, M. R. Stinson, and D. I. Havelock, "Experiments on surface waves over a model impedance plane using acoustical pulses," J. Acoust. Soc. Am. **99**, 1993–2005 (1996).
[25]K. Attenborough, "Ground parameter information for propagation modeling," J. Acoust. Soc. Am. **92**, 418–427 (1992).
[26]D. D. Rife and J. Van der Kooy, "Transfer-function measurement with maximum-length sequences," J. Audio Eng. Soc. **37**, 419–443 (1989).
[27]T. A. Busch and M. R. Hodgson, "Improved method for selecting scale factors and model materials for scale modelling of outdoor sound propagation," J. Sound Vib. **243**, 173–181 (2001).
[28]K. V. Horoshenkov, D. C. Hothersall, and S. E. Mercy, "Scale modelling of sound propagation in a city street canyon," J. Sound Vib. **223**, 795–819 (1999).

A model for the perception of environmental sound based on notice-events

Bert De Coensel[a)]
Department of Electrical Engineering and Computer Sciences, Computer Science Division, University of California at Berkeley, Berkeley, California 94720-1776

Dick Botteldooren and Tom De Muer
Department of Information Technology, Acoustics Group, Ghent University, B-9000 Ghent, Belgium

Birgitta Berglund and Mats E. Nilsson
Gösta Ekman Laboratory for Sensory Research, Karolinska Institutet and Stockholm University, SE-10691 Stockholm, Sweden

Peter Lercher
Division of Social Medicine, Medical University Innsbruck, A-6020 Innsbruck, Austria

(Received 18 May 2008; revised 13 April 2009; accepted 13 May 2009)

An approach is proposed to shed light on the mechanisms underlying human perception of environmental sound that intrudes in everyday living. Most research on exposure-effect relationships aims at relating overall effects to overall exposure indicators in an epidemiological fashion, without including available knowledge on the possible underlying mechanisms. Here, it is proposed to start from available knowledge on audition and perception to construct a computational framework for the effect of environmental sound on individuals. Obviously, at the individual level additional mechanisms (inter-sensory, attentional, cognitive, emotional) play a role in the perception of environmental sound. As a first step, current knowledge is made explicit by building a model mimicking some aspects of human auditory perception. This model is grounded in the hypothesis that long-term perception of environmental sound is determined primarily by short notice-events. The applicability of the notice-event model is illustrated by simulating a synthetic population exposed to typical Flemish environmental noise. From these simulation results, it is demonstrated that the notice-event model is able to mimic the differences between the annoyance caused by road traffic noise exposure and railway traffic noise exposure that are also observed empirically in other studies and thus could provide an explanation for these differences.
 [DOI: 10.1121/1.3158601]

PACS number(s): 43.50.Rq, 43.50.Qp, 43.66.Lj, 43.66.Ba [BSF] Pages: 656–665

I. INTRODUCTION

During the last century, substantial research efforts have been spent on relating noise exposure to the discomfort or annoyance caused by it. A large body of knowledge is based on socio-acoustical surveys, in which large groups of persons are questioned about their own home environment. The annoyance effects of noise and potential confounders are mainly expressed on the basis of population averages.[1,2] Researchers have, for example, focused strongly on determining quantitative relationships between community noise annoyance of transportation noise and outdoor energy equivalent sound pressure levels [L_{dn} or L_{den} (Refs. 3 and 4)]. In practice, noise annoyance modifiers have been modeled by adding penalties or bonuses to the equivalent sound pressure levels.[5,6]

It has been suggested that sound events may play an important role in the prevalence of noise annoyance.[7–12] This can be argued based on the principle that the auditory system, like the other sensory systems, records instant changes in environmental stimuli rather than the average level of a continuous exposure.[10] Various indicators have been used for characterizing noise as a series of sound events, such as the maximum sound pressure level L_{Amax}, the number of (heavy) vehicle pass-bys, or the number of events for which the sound pressure level exceeds a given threshold.[13–15] These are alternative exposure models for expressing the relationship with annoyance.[16]

Several authors have put forward the assumption that sound has to be noticed in order for it to contribute to an overall impression of annoyance.[17–21] Noticing depends on the characteristics of the sound (the most important being the signal-to-noise ratio) and on the observer listening and paying attention to the sound. The indicators do neither include this continuously active attention process nor account for the necessary co-occurrence of the listening process of the continuous sound. In this work we introduce the concept of a *notice-event*, which is defined as an instance of consciously perceiving the sound under study. We start from the hypothesis that long-term perception of environmental sound (and

[a)]Author to whom correspondence should be addressed. Electronic mail: bert.decoensel@intec.ugent.be. On leave from Acoustics Group, Department of Information Technology, Ghent University, Belgium.

reported annoyance) is built from a series of such notice-events and their sound characteristics during these notice-events. To estimate notice-events and thus to unravel part of the underlying mechanisms of human perception of environmental sound, we propose a computational *ab initio* approach. To illustrate this concept, we will develop a causal simulation model for auditory perception, based on known acoustical, psychoacoustical, and psychological principles, extracted from well-controlled experiments on perception reported in literature.

In Sec. II, we will outline the proposed methodology from a conceptual point of view. In Sec. III, a quantification and implementation of some of the introduced concepts will be discussed. In Sec. IV, we will use the proposed model to explain differences in traffic noise annoyance.

II. METHODOLOGICAL FRAMEWORK

A. Definition of a notice-event

Our key hypothesis states that *the perception of environmental sound is primarily determined by consciously noticed sounds*. The importance of noticing in the perception of complex sound has already been stressed in earlier work concerning the emergence of noise annoyance.[17–21] It could be argued that the "non-noticed" background sound could affect mood, stress, and certainly health. However, in light of the application envisaged in this paper such higher-order effects are assumed to be of minor importance and are therefore not considered.

We define a notice-event as an instant of consciously perceiving the sound under study. The occurrence of a notice-event depends, among other things, on the sound level of the sound relative to the background, the current environmental context of the listener and his/her activity (which may also produce sound, if, e.g., watching television or cooking), as well as on several personal factors which vary with time, such as the degree of focused attention to the sound[22] and the amount of (short-term) mental habituation[23,24] that may have occurred. It is important to note that a notice-event can occur whenever any of the above-mentioned parameters changes. For example, a sudden increase in attention may trigger a notice-event, even though the sound level did not increase. Instantaneous attention to environmental sound will be influenced by changes in overall alertness toward ones environment with time (e.g., circadian variation[25]) and will depend on the ongoing activity. For example, sleeping lowers the conscious attention to sound, whereas relaxing may result in heightened attention to sound. While performing a task that requires significant mental resources, attention to sound is likely to be low.[19]

B. *Ab initio* modeling

In order to estimate the occurrence of notice-events, we propose to use a computational *ab initio* approach. The increasing availability of computing power has made it possible to simulate complex systems *ab initio*, starting from basic laws governing the behavior of particular constituents. Mainly because of its predictive capacity and its potential for discovering new phenomena, *ab initio* modeling has become very popular in natural as well as applied sciences.[26,27] Next to predicting system behavior, computational *ab initio* models have great potential in extracting knowledge on the internal mechanisms governing this behavior. For example, the frontiers of computing are now explored to build computational models of psychological processes such as visual perception.[28] By implementing a model in software, one is forced to concretize fuzzy ideas, to quantify model parameters, or even to formulate hypotheses and to add additional submodels. Constructing the model in itself is therefore a fruitful process since it forces the scientist to make explicit the knowledge on the mechanisms governing the behavior of the system. The model is called valid if it is able to recreate, at least in a qualitative way, macroscopic trends and dependencies found in empirical data from the real-life system. The mathematical description of the model then arises gradually by iteration between software implementation, simulation, and comparison with real-life phenomena.

In applying an *ab initio* approach to the estimation of the occurrence of notice-events, the aim is to build a logical model for an individual person based on known acoustical, psychoacoustical, and psychological principles. A model describing an individual person will have to include the auditory perception of environmental sound in the everyday living environment, together with its underlying mechanisms in a broad sense. Furthermore, an *ab initio* approach requires a *causal* model for perception. Inclusion of causality naturally leads to the inclusion of temporal aspects, making perception emerge and transform over time. Explicitly accounting for time also provides opportunities to model the interaction of factors influencing perception and its evaluation over time. Consider, for example (short-term) habituation to noise, which may suppress the occurrence of notice-events over time. A time-domain model will allow linking this basic understanding of the mechanisms to the gross effect observed in surveys.

An often used paradigm is to model the internal mechanisms governing the behavior of individuals, and then to observe the group behavior resulting from a large number of simulated individuals (sometimes referred to as a *synthetic population*), each characterized by its own set of parameters to model interindividual differences. Exposure-effect relationships valid for large groups of persons can also be extracted from averaging over a synthetic population, but in the case of environmental sound perception, the interactions between group members have low influence on the overall behavior. Interindividual interactions are ignored in the simplified model presented in this paper.

III. A MODEL FOR NOTICE-EVENTS

A. General considerations

To substantiate the ideas introduced in Sec. II, an auditory perception model for estimating the occurrence of notice-events is developed mathematically in this section. A time-domain model is envisaged because, as already shown above, temporal effects need special care. Ranges for time constants are derived from the psychoacoustical literature

and from a specific noise annoyance field experiment.[29] A slightly different version of the model has earlier been reported elsewhere.[30–33]

In view of the feasibility of the numerical effort, details of the auditory processing of the sound signal are not considered: time-varying descriptors of what is potentially perceived by the auditory system form the input of the model, rather than the raw sound wave. Time sequences of the sound pressure level from typical sources such as cars, trains, birds, etc., are simulated based on the presence of sources in the environment (Sec. III E). If one is situated in an environment with multiple sound components, the acoustical pattern at the ear will consist of the sum of all concurrent sounds. Nevertheless, the human auditory system is able to separate this mixture of sounds[34] and to form separate descriptions of each sound source. This mechanism is commonly referred to as *auditory scene analysis*.[35] By simulating auditory streams for each separate environmental sound source, the non-trivial problems of modeling auditory scene analysis and sound source identification[36] are by-passed. The difference between mere detecting and identifying is thereby ignored. In other words, it is assumed that a non-identifiable sound is not noticed. Note that also the effects of tonality—the contribution to source identification is implicit—are neglected since only the sound pressure level is considered.

B. Noticed sound exposure level

The model for auditory perception analyzes the environmental sound and identifies its noticed sound components. The difference in sound pressure level between any particular sound and the "background" noise (all other sounds) is considered as the key feature; once this difference exceeds a (dynamic) threshold, a notice-event occurs. The condition used to identify the starting point in time of a notice-event is simplified to

$$L_f(t) > L_b(t) + T_{\text{start}}(t), \tag{1}$$

with t the current simulation time, L_f the sound pressure level of the sound under study (foreground sound), L_b the background sound pressure level, and T_{start} the notice-threshold [sound pressure level in dB(A)] for the start of a notice-event. The instantaneous notice-threshold can be interpreted as the signal-to-noise ratio that the foreground "signal" sound has to have to be noticed consciously above the background "noise." In a similar way, the point in time at which the modeled individual stops in noticing the foreground sound is determined by

$$L_f(t) < L_b(t) + T_{\text{stop}}(t), \tag{2}$$

with T_{stop} the notice-threshold for the end of a notice-event. Together, T_{start} and T_{stop} mark segments $[t_b^i, t_e^i]$ in time at which the sound under study is noticed.

Inspired by the hypotheses that only consciously noticed sound events contribute to annoyance, and that the contribution of an event to perceived annoyance is proportional to its audibility above the background,[20] we define the *noticed sound exposure level* by

$$\text{SEL}_{\text{thr}} = 10 \log_{10}\left(\sum_{i=1}^{N} \int_{t_b^i}^{t_e^i} 10^{L_{\text{thr}}(t)/10} dt\right), \tag{3}$$

where N denotes the number of notice-events during the simulated time period, t_b^i and t_e^i denote the beginning and end times of these notice-events, and L_{thr} denotes the portion of the sound pressure level during a notice-event that is above the notice-threshold:

$$L_{\text{thr}}(t) = L_f(t) - L_b(t) - T_{\text{stop}}(t). \tag{4}$$

The idea of only taking into account the contribution of sounds above a threshold in the calculation of the sound exposure level has in the past been proposed to be a more logical way to assess noise annoyance.[18,37] Instead of using a fixed threshold, this work considers the threshold to be time-varying and depending on the instantaneous attention to the specific component sound.

C. Attention and gating

The degree of attention to a specific sound source is dependent on a number of factors. As already mentioned, attention will depend on the time of the day and on the current activity. It is safe to assume that $T(t)$ decreases monotonically with increasing attention $a(t)$; hence for simplicity, we approximate it by a linear function

$$T'_{\text{start}}(t) = T_{\text{stop}}(t) = T_{\text{base}} - C_n \cdot a(t), \tag{5}$$

where C_n is used as a general constant determining the influence of attention on the notice-threshold.

After a sound is noticed, subsequent peaks in its sound pressure level will not trigger the beginning of a new notice-event, unless these are sufficiently more noticeable. The psychophysical function closest to explaining this is *gating* (perceptual and attentional origin). The gating condition holds on slightly longer than the event that was first noticed continues because it is known that non-negligible time constants would be involved in the process.[38,39] Hence, Eq. (5) is extended to model gating:

$$T_{\text{start}}(t) = T'_{\text{start}}(t) + C_g \exp\left(-\mathcal{H}(t-\tilde{t}_e) \cdot \frac{t-\tilde{t}_e}{\tau_g}\right), \tag{6}$$

where $\tilde{t}_e$ is the end time of the last notice-event ($+\infty$ if t lies within a notice-event). $\mathcal{H}$ denotes the Heaviside step function. As a result of gating, multiple peaks in sound pressure level occurring shortly after each other are lumped into one notice-event. C_g determines the extra sound level, relative to the level of the current sound event, that notice-events need in order to defeat the gating mechanism. The time constant τ_g determines the decrease in noticeability immediately after a notice-event. This mechanism determines whether, for example, individual cars and trucks are noticed, or rather the traffic as a whole.

D. Habituation and attention focusing

Because of (short-term) mental habituation,[23,24,40] the effective duration of noticing a sound can be shorter than what is expected from the raw sound pressure levels. Habituation

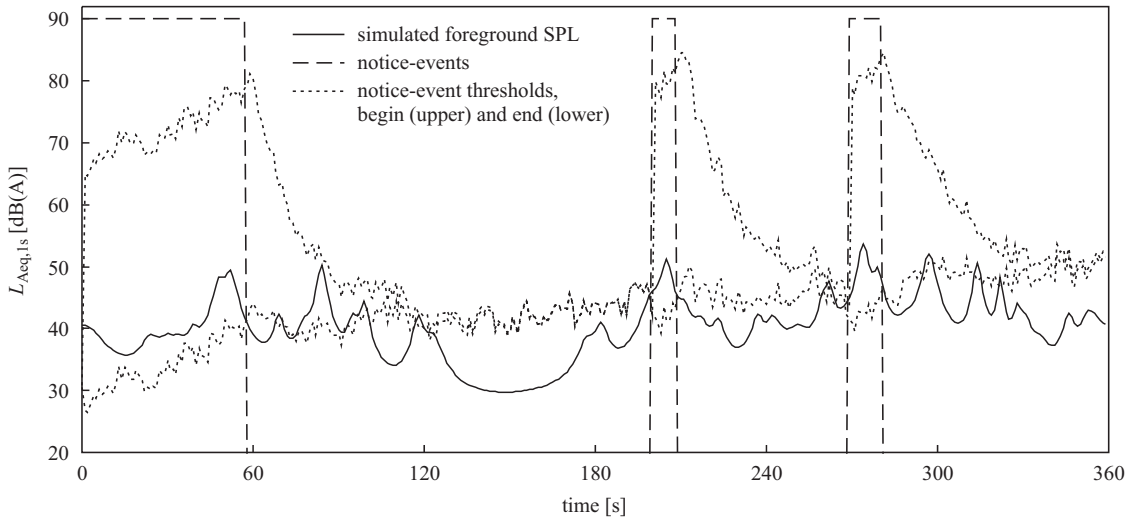


FIG. 1. Excerpt of a simulated foreground $L_{Aeq,1\,s}$ time series $L_f(t)$ (solid lines) together with the notice-events (delimited time periods in dashed lines) and the corresponding thresholds for beginning $[L_b(t)+T_{start}(t)$, upper dots] and end $[L_b(t)+T_{stop}(t)$, lower dots)] of a notice-event.

may also suppress the emergence of new notice-events because the individual has become used to a certain sound exposure. However, it is well-known that habituation is limited.[23,24] Habituation $h(t)$ is assumed to be proportional to the exponentially averaged sound pressure level,[38,41] whether the sound is noticed or not,

$$h(t) = \frac{1}{\tau_h}\int_{-\infty}^{t} L_f(u)\cdot \exp\left(-\frac{t-u}{\tau_h}\right)du. \tag{7}$$

Time constants τ_h involved in stimulus specific short-term habituation vary between several seconds to several minutes.[38,41] Furthermore, it is assumed that the effect of short-term habituation is mediated through reduced attention to intruding sound,

$$a'(t) = a_{act}(t) - C_h \cdot h(t), \tag{8}$$

where $a_{act}(t)$ denotes the (potentially time-varying) attention to the particular sound associated with the current activity, and where C_h determines the strength of the habituation effect.

When a modeled individual notices a sound, by definition, his/her attention is drawn to it. Because of this elevated state of attention, events in the immediate future are more likely to be noticed. This effect counteracts habituation but on a shorter time scale (a couple of seconds). It also reduces the effect of gating but at a different level: attention focusing is part of a cognitive process, whereas gating is sensory in nature. Formally, we may define attention focusing as a priority shift that a notice-event induces in the global attention system, resulting in a temporary increase in attending to sound.

Attention focusing is instantaneous and diminishes over time. When no new notice-events occur, the priorities reorganize and fall back to their equilibrium. Attention focusing can be modeled by a step in the attention that exponentially fades away:

$$a(t) = a'(t) + C_a \exp\left(-\frac{t-\tilde{t}_b}{\tau_a}\right), \tag{9}$$

with $\tilde{t}_b$ the starting time of the last notice-event and τ_a the time constant involved in attention focusing. Note that the product $C_n \cdot C_a$ determines the reduction in notice-threshold caused by the attention focusing of the modeled individual.

Figure 1 shows an example of a time series of simulated foreground sound pressure level caused by road traffic (see Sec. III E), together with the notice-events. The simulation starts when the modeled individual is suddenly exposed to road traffic sound, for example, because he or she leaves the home. At that moment, road traffic is noticed, but the modeled individual becomes habituated to the sound (indicated by an elevated threshold) and it will not be noticed anymore after about a minute. Because of the gating and habituation mechanisms, vehicles passing by during the subsequent minutes will not be noticed until a significantly louder vehicle passes by after approximately 3 min and 20 s.

E. Simulating sound exposure

The first design decision is to choose an appropriate temporal resolution for the simulation based on the expected resolution of notice-events. The time to react to a sound is around 0.2 s for the sound pressure levels considered here[42] (well above the hearing threshold). However, the psychological presence, that is, the perceived *nowness*, has been estimated to be 2–3 s.[43,44] Peaks in the sound pressure level would be smoothened when using such a large temporal resolution. Therefore, 1 s is chosen as the temporal resolution: a safe margin on the perception side and sufficiently small to resolve individual notice-events. The input to the model for auditory perception will be expressed as a time series of $L_{Aeq,1\,s}$ values for each of the environmental sounds considered.

Various sound sources can be simulated to reproduce the environmental sound to which a modeled individual is exposed. In light of Sec. IV, in which our model will be used to

predict transportation noise annoyance in an at-home context, we will first focus on the noise produced by road traffic and railway traffic. The generation of the fluctuations in sound pressure level caused by transportation could be accomplished by noise mapping software capable of calculating time-varying noise immission.[45,46] Generally, this entails simulating road traffic and railway traffic noise in the large neighborhood of the modeled individual, calculating the time-varying noise emission associated with all vehicle sources and finally calculating the resulting noise immission at the location of the modeled individual by the use of a sound propagation model. However, this approach is computationally very intensive and thus only feasible for small areas.

For the example given in this paper, it is assumed that the road traffic noise exposure of a modeled individual is mainly determined by the main road closest to its location. This road is modeled as two single lane vehicle streams, one of passenger vehicles and one of heavy vehicles. In its current form, the model does not include binaural effects and thus the exact location and driving direction of vehicles would be irrelevant. The calculation of attenuation is simplified by assuming a free field and by ignoring the presence of screening obstacles and meteorological effects.

For the example given in this paper, we also simplify traffic statistics. The duration between road vehicle passages is primarily determined by individual driver characteristics. In considering a stream of vehicles as a whole, the time between two consecutive vehicle passages can be modeled in first order approximation by an exponential distribution. This implies that the occurrence of a vehicle passage during a sufficiently small time interval can be modeled by a Poisson process:[12,47] the probability of a vehicle passage only depends on the traffic intensity and is independent of earlier passages. Furthermore, vehicle speed is found to adhere to a normal distribution.[48] These approximations will fail when the road is close to saturation (e.g., during rush hour). However, at that point the sound pressure level produced by the road is more invariant and notice-events will be determined mainly by other changes (e.g., a change in activity).

Road vehicles are modeled by an omnidirectional point source. The A-weighted squared sound pressure at the location of the modeled individual (before insulation of the building is accounted for), produced by a single vehicle, is approximated by

$$p^2(t) = \frac{1}{4\pi} \cdot \frac{W \cdot \rho c}{[v \cdot (t - t_p)]^2 + d_{\text{road}}^2}, \tag{10}$$

where ρc denotes the characteristic impedance of air, W denotes the A-weighted sound power of the vehicle, t_p denotes the time of occurrence of the vehicle passage, and v denotes the vehicle traveling speed on a straight road at perpendicular distance d_{road} to the modeled individual. A vehicle stream is characterized by the traffic intensity and the mean and variance of the vehicle speed. Using this technique to predict the noise exposure allows us to include easily some variability of the speed of the vehicles within a stream (and consequently their noise-source power) when modeling the observed distribution of emissions and speeds in real life. Finally, the noise produced on the nearest main road by both passenger vehicle and heavy vehicle streams is added together to obtain the total instantaneous sound exposure produced by road traffic noise.

Similar to the case of road traffic noise, it is assumed that the railway noise exposure of a modeled individual is mainly determined by the railway closest to its location. This railway noise is modeled as two train streams, one of person trains and one of freight trains. Although railway traffic is much more deterministic than road traffic, the same approach for pass-bys is followed: arrival times are assumed to adhere to a Poisson distribution. The difference is that trains are modeled as moving line sources rather than as moving point sources. The A-weighted squared sound pressure at the location of the modeled perceiving individual (before insulation of the building is accounted for), produced by a single train, is approximated by

$$p^2(t) = \frac{1}{4\pi \cdot M} \sum_{i=0}^{M} \frac{W \cdot \rho c}{\left[v \cdot (t - t_p) + \frac{i \cdot l}{M}\right]^2 + d_{\text{rail}}^2}, \tag{11}$$

where l denotes the length of the train and M denotes the number of point sources used to represent the sound from the line source; typically M is around 20.

Next to the noise caused by transportation, the sound to which the modeled individual is exposed in its dwelling may consist of, among other things, the sound generated by the individual itself and possibly by other individuals, various mechanical and electrical sounds, and sounds from nature (wind, birds, etc.). It is not feasible to model accurately all these sources of sound. However, because we are only interested in the temporal envelope of the sound produced by the above-mentioned sources, we can take advantage of known statistical properties of the temporal envelope of ambient sounds.

The spectral density of the temporal envelope of various types of music and speech has often been found to show a remarkably good resemblance to $1/f$ noise[49] on time scales spanning minutes to hours. Furthermore, this $1/f$ temporal behavior has been revealed in a broad range of environmental sounds, both in rural and urban contexts. This behavior can be linked to self-organization of the underlying complex systems of noise sources.[50,51] Based on these findings, we decided to model the temporal envelope of the sound pressure level caused by all sources other than transportation as $1/f$ noise. It could be argued that the sound produced by the modeled individual itself should be simulated separately because the modeled individual is the producer and has some form of control and because this sound is correlated with attention through activity. However, because the model presented here does not at present include behavioral feedback mechanisms, such as active coping, a separation would have little influence on the outcome.

TABLE I. Range for the parameters of the modeled individuals in the synthetic population.

Parameter	Description	Interval	Unit
T_{base}	Base threshold	0–20	dB(A)
C_n	Attention strength	0–20	dB(A)
a_{act}	Activity attention	0–3	–
C_g	Gating strength	0–50	dB(A)
τ_g	Gating time constant	0.1–60	s
C_h	Habituation strength	0–0.2	dB(A)^{-1}
τ_h	Habituation time constant	0.1–200	s
C_a	Focusing strength	0–1	–
τ_a	Focusing time constant	0.1–10	s

IV. SIMULATION RESULTS

A. Sensitivity analysis

The model for auditory perception presented in Sec. III contains a number of parameters. Although important knowledge is available from well-controlled perception experiments, exact values for all parameters are not readily available in the literature, especially not if the extent of interindividual differences are to be fully considered; some parameters are hard or impossible to measure. As mentioned in Sec. II B, one solution is to consider a synthetic population consisting of a large number of individuals and to draw samples from appropriate distributions for each parameter instead of assigning fixed values. Ideally, the synthetic population will exhibit the same statistics on the observed variables as would be found in observations of the real population. As a first numerical illustration, we will investigate the sensitivity of the auditory perception model to each of its parameters using this sampling-based technique. The intention is to find those model parameters which distributions are most responsible for the variation in model output.

Let us consider a synthetic population of 10^4 individuals with model parameters that are randomly sampled from the intervals shown in Table I. Because of lack of data, a uniform distribution over individuals is assumed for all parameters. Intervals were chosen wide enough to encompass the ranges found in the literature cited in Sec. III. The following fixed contextual parameters are considered: each individual is located inside its dwelling, which has an insulation of 20 dB(A); the average ambient sound pressure level (including the noise produced by the individual itself) is 40 dB(A); a road is located outside the dwelling at a distance of 20 m, carrying a flow of 600 vehicles/h with 20% of heavy vehicles, and with an average vehicle speed of 70 ± 3 km/h. During daytime, these values are typical for home environments in Flanders, the northern part of Belgium (see also Sec. IV B).

For each modeled individual, time series of $L_{Aeq,1\,s}$ inside the dwelling caused by road traffic and other ambient noise sources were simulated for a duration of 1 h, using the model described in Sec. III E. To estimate the *A*-weighted sound power of the road vehicles as a function of vehicle type and speed, the Harmonoise model[52] was used with reference conditions for road surface. Subsequently, a series of notice-events was calculated for each modeled individual, considering the road traffic noise as the foreground sound and the $1/f$ ambient noise as the background sound.

Figure 2 shows the noticed sound exposure level SEL_{thr} [Eq. (3)] by the simulated individuals of the synthetic population as a function of each model parameter (averaged within ten sub-intervals). It can be seen that the base notice-threshold T_{base} and the habituation strength C_h have the largest influence on SEL_{thr}. As would be expected, less vehicle passages will be noticed if the base notice-threshold is high, if the habituation time constant is low, or if the influence of habituation is high. As a consequence, changes in sampling distribution for these parameters will have the largest influence on aggregated results. Gating and focusing seem to have little influence if results are averaged over all other model parameters. However, the model contains several interactions between parameters, e.g., results are more sensitive to a_{act} if C_n is high [this can be seen from Eqs. (5), (8), and (9)] and to τ_g if C_g is high [Eq. (6)]. As a consequence, changes in the sampling distribution for these parameters also could have some influence, although this influence will be smaller than comparable changes in the sampling distribution for T_{base} or C_h. The relationships shown in Fig. 2 still hold qualitatively if traffic flow, speed, and distance to the closest main road are varied within realistic bounds.

B. Representative context

As a second numerical illustration of the model presented and discussed in Sec. III, we consider a representative sample of the population in Flanders. In order to achieve a set of contexts, representative for home environments in Flanders, a synthetic population was constructed to resemble the population living at 7500 sampled locations. The dwellings were located geographically using their addresses, and these were linked to Geographic Information System (GIS) data (road segments and traffic data). Subsequently, the perpendicular distance to the nearest main road was calculated for each dwelling, as was the average hourly intensity and speed (average value and standard deviation) of cars and heavy vehicles during daytime, evening, and nighttime. The simulation parameters of the railway closest to the dwelling were determined by the same algorithm as for the road.

Time series of $L_{Aeq,1\,s}$ levels inside the dwelling caused by road traffic, railway traffic, and other ambient noise sources were simulated for a duration of 24 h. For road traffic, again the Harmonoise model[52] was used. To estimate the *A*-weighted sound power of the trains as a function of train type and speed, the BeGIS model[53] was used. This model combines measurements of the local rolling stock with trends from the Dutch standard method. Because of lack of empirical data, the average ambient sound pressure level and the building insulation were each randomized among individuals. The distribution of physical noise insulation of building façades behind which people reside most of the day ranged from 0 (e.g., individual in garden) to 30 dB(A) during daytime and evening and from 10 to 30 dB(A) during nighttime. The average ambient sound pressure level, including the noise produced by the modeled individual's activities, ranged from 25 to 55 dB(A) during daytime and evening and from

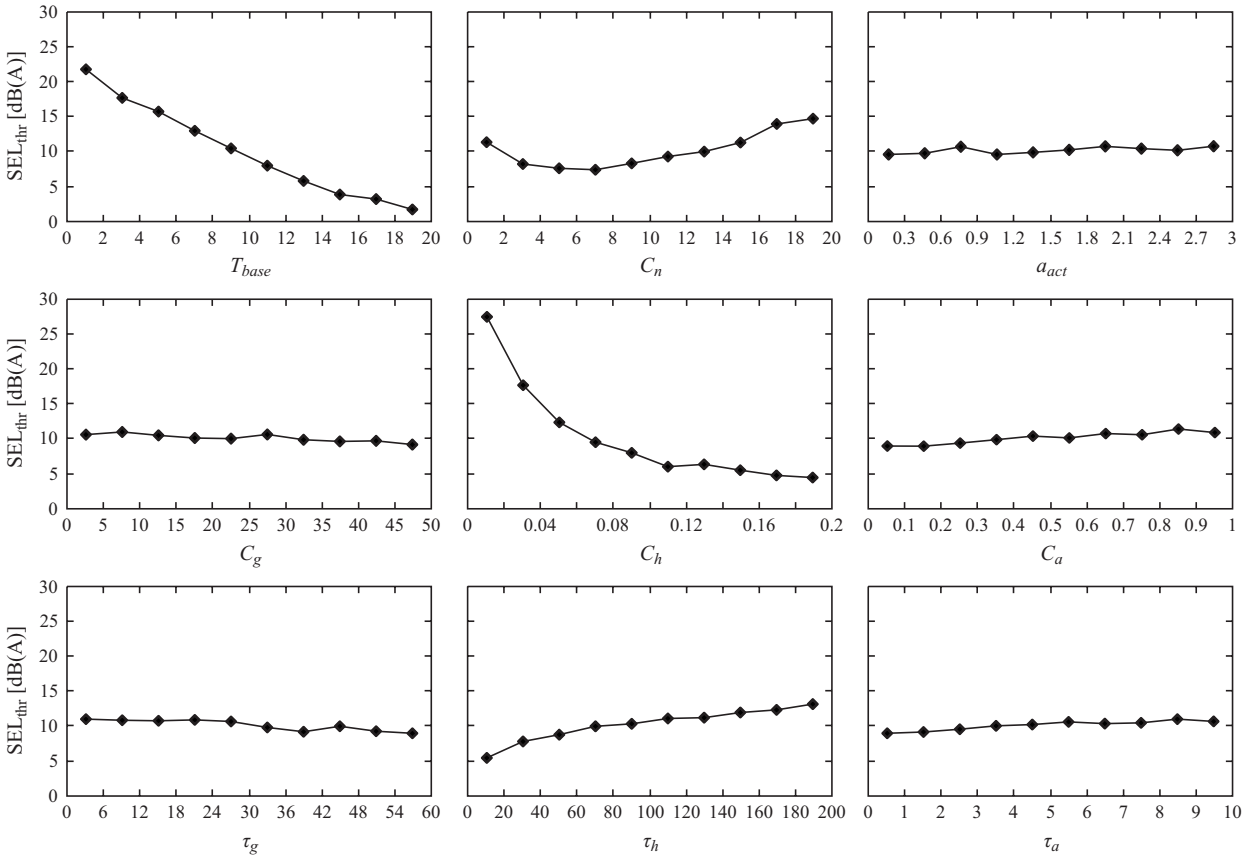


FIG. 2. Noticed sound exposure level for 1 h for the simulated individuals of the synthetic population caused by road traffic noise exposure as a function of the different model parameters.

15 to 35 dB(A) during nighttime. Model parameters were again randomly sampled from the intervals shown in Table I.

Two series of notice-events were calculated for each modeled individual:

(1) considering the major road traffic noise as the foreground sound, and the combination of railway traffic noise and $1/f$ ambient noise as the background sound and
(2) considering the railway traffic noise as the foreground sound, and the combination of major road traffic noise and $1/f$ ambient noise as the background sound.

Figure 3 shows the noticed sound exposure level by the simulated individuals of the synthetic population, as a function of façade L_{den}, for road and railway traffic. Above an L_{den} of 45 dB(A), a distinct difference of $\mathrm{SEL}_{\mathrm{thr}}$ is evident between road and railway traffic noise at the same average sound pressure level. This difference resembles remarkably differences reported in annoyance between road and railway traffic noise, as observed in several field studies of the past.[4,54,55] These annoyance differences have led to less restrictive regulation for railway noise (railway bonus) relative to noise from road traffic (see, e.g., the German, French, or Austrian legislation). Several explanations for a railway bonus have been proposed.[4,29]

How can the difference in annoyance for road and railway traffic be explained in light of the model of Sec. III? Consider the sound event caused by a train passage. This event will be noticed if its sound pressure level is sufficiently high at the same time as the levels of the background noise and the notice-threshold are sufficiently low [Eq. (1)]. With a fluctuating sound pressure level of the background and of the notice-threshold, the duration of the event plays a role. If the sound pressure level of the train event is not too high, by chance the sound pressure level of the background and of the notice-threshold have to be low *during* the train passage. Otherwise the train event will not be noticed. Therefore, the chance to notice major road traffic of invariant sound pressure level will be higher (although lower than for the maximum sound pressure level of the train passage at the same L_{Aeq}).

If train noise is considered on its own, it is expected that habituation will gradually decrease the probability of noticing, whereas attention may increase this probability. If the noise from the major road is added to this scenario, it will act as an additional background noise to the noticed train passages, in particular, if sound pressure levels are sufficiently low. The reduced signal-to-noise ratio will decrease the overall probability of noticing. However, this mechanism is not reciprocal: train noise events, because of their relatively lower rate of occurrence, are expected to influence less the noticing of the noise from the major road. Also, because road traffic noise has a more continuous character, habituation will influence largely the notice-threshold for road traffic noise.

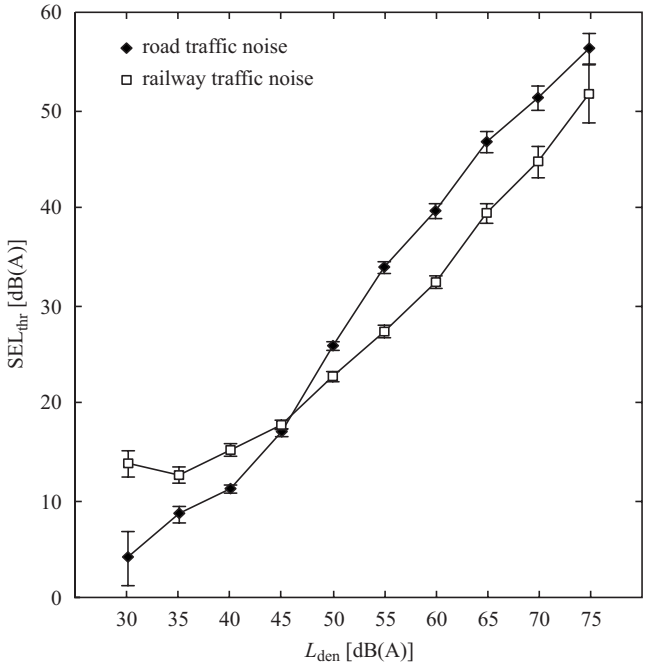


FIG. 3. Noticed sound exposure level (24 h) for the simulated individuals of the synthetic population caused by road (◆) and railway (□) traffic noise exposure as a function of L_{den} caused by the respective sources. The standard error is marked.

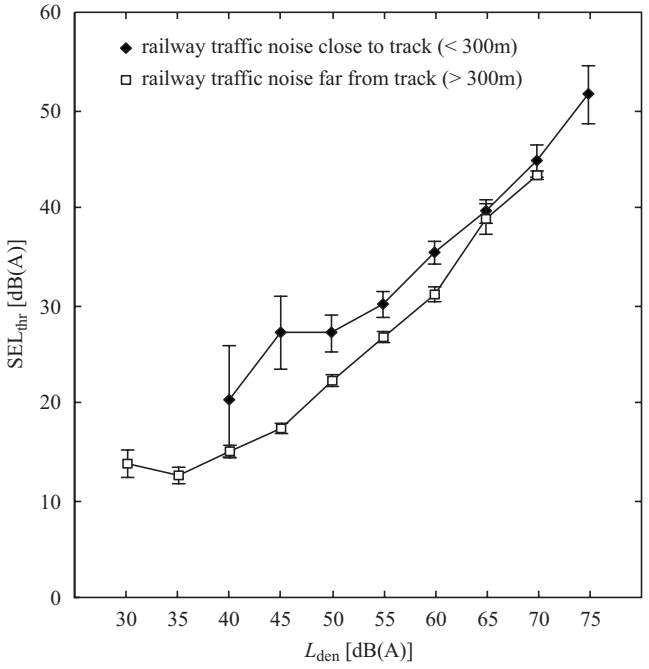


FIG. 4. Noticed sound exposure level (24 h) for the simulated individuals of the synthetic population caused by railway traffic noise exposure as a function of L_{den} for railway tracks close to the dwelling (◆, <300 m) and far from the dwelling (□, >300 m). The standard error is marked.

Figure 4 shows the noticed sound exposure level by the simulated individuals of the synthetic population, caused by railway traffic noise exposure, plotted as a function of calculated railway L_{den}. The data are grouped according to distance to the railway track. A railway track at a large distance will, for the same average sound pressure level, produce lower maximum levels because of the spread over time of the sound pressure level envelope of the passages. As a consequence, the energetic masking accomplished by road traffic noise and other ambient noise sources is expected to be larger. This would explain the lower probability of noticing a train passage at a large distance, as compared to a train passage at a close distance. Similar to this observation, it has previously been shown that annoyance to railway traffic noise depends significantly on the distance to the track.[29,56,57]

V. CONCLUSIONS AND PERSPECTIVES

A theoretical modeling approach to the perception of environmental sound in the home environment has been presented. Our proposal was to model, in a bottom-up fashion, the perception of an individual *a priori* from known acoustical, psychoacoustical, and psychological knowledge. Ideally, this approach will take into account personal and contextual variables in clarifying the human perception of environmental sound. To outline the idea, a computational model for auditory perception was constructed, by which psychoacoustical processes such as gating, habituation, and attention focusing were implemented.

The presented model contained a number of parameters, but because results are to be considered only at an aggregated level, no fixed values need to be assigned to these parameters. Instead, values can be sampled from appropriate distributions, with wide enough intervals to encompass the ranges found in literature. As a consequence, part of the variability in model output results from the variability introduced through sampling model parameters. A sensitivity analysis showed that the base notice-threshold and the habituation strength have the largest influence on the variability of the model output, provided results are averaged over all other model parameters. These parameters could constitute a means for including the effects of, e.g., the current activity or sensitivity to noise, but more research will be needed to establish the appropriate relationships.

As an illustration of the model, simulation results were presented based on a representative sample of the population in Flanders. The noticed sound exposure level was defined to measure the fraction of the sound exposure that is consciously noticed. If one accepts that noticing an environmental sound is a prerequisite for becoming annoyed (defined here as self-reported annoyance in surveys), the proposed notice-event model allows us to *explain* two features observed in the literature of the exposure-effect relationships for self-reported noise annoyance and L_{den}.

First, in the living environment, railway traffic noise is found to be less annoying than road traffic noise at the same average sound pressure level.[4,54,55] The proposed model suggests that this difference is based on physical properties and basic psychophysical processes of the human auditory system. Differences in cognitive processing and emotions were not included in the model and thus seem of less importance for explaining the effect—as has sometimes been suggested. Accordingly, the difference between railway traffic and road traffic noise annoyance is mainly caused by the temporal structure of the exposure, by noticing, and by associated attention and habituation processes.

Second, in many studies, increases in annoyance were observed to be caused by railway noise at shorter distance to the track although the L_{den} was the same, though this effect is not documented as consistently in the literature as the railway bonus. Indeed, for the exposure situation in Flanders, the notice-event model shows a higher noticed sound exposure level at shorter distances. An important factor in our notice-event model is the background noise, e.g., road traffic noise and noise from own activities in the home, which decides how much the noise maximum of the train passage emerges above it. Trains at shorter distance produce higher peak levels which emerge more easily above this background, even at low average levels.

Many authors have pointed out the importance of personal factors, including personality traits and states, in the assessment of environmental sound.[1,5,6,58,59] Our approach treats each individual as being unique and tries to model contextual and personal factors on an individual basis. Therefore, future developments of the proposed model should include the implementation of emotional and cognitive submodels, which should allow for further effects on annoyance as regards personal preference, the built environment, living circumstances, and the socio-cultural context.[60]

ACKNOWLEDGMENTS

B.D.C. is a postdoctoral fellow of the Research Foundation-Flanders (FWO-Vlaanderen) and of the Belgian American Educational Foundation; the support of both organizations is gratefully acknowledged. Furthermore, this research was partially sponsored by grants from the Swedish Foundation for Strategic Environmental Research (MISTRA), the Swedish National Road Administration, and the Swedish Agency for Innovation Systems (Vinnova).

[1]R. F. S. Job, "Community response to noise: A review of factors influencing the relationship between noise exposure and reaction," J. Acoust. Soc. Am. **83**, 991–1001 (1988).
[2]P. J. M. Stallen, "A theoretical framework for environmental noise annoyance," Noise Health **1**, 69–79 (1999).
[3]T. J. Schultz, "Synthesis of social surveys on noise annoyance," J. Acoust. Soc. Am. **64**, 377–405 (1978).
[4]H. M. E. Miedema and H. Vos, "Exposure-response relationships for transportation noise," J. Acoust. Soc. Am. **104**, 3432–3445 (1998).
[5]H. M. E. Miedema and H. Vos, "Demographic and attitudinal factors that modify annoyance from transportation noise," J. Acoust. Soc. Am. **105**, 3336–3344 (1999).
[6]H. M. E. Miedema and H. Vos, "Noise sensitivity and reactions to noise and other environmental conditions," J. Acoust. Soc. Am. **113**, 1492–1504 (2003).
[7]R. Rylander, S. Sörensen, and A. Kajland, "Traffic noise exposure and annoyance reactions," J. Sound Vib. **47**, 237–242 (1976).
[8]R. Rylander, M. Björkman, U. Åhrlin, S. Sörensen, and K. Berglund, "Aircraft noise annoyance contours: Importance of overflight frequency and noise level," J. Sound Vib. **69**, 583–595 (1980).
[9]S. Sörensen and N. Hammar, "Annoyance reactions due to railway noise," J. Sound Vib. **87**, 315–319 (1983).
[10]M. Björkman, "Community noise annoyance: Importance of noise levels and the number of noise events," J. Sound Vib. **151**, 497–503 (1991).
[11]T. Sato, T. Yano, M. Björkman, and R. Rylander, "Road traffic noise annoyance in relation to average noise level, number of events and maximum noise level," J. Sound Vib. **223**, 775–784 (1999).
[12]M. J. Roberts, A. W. Western, and M. J. Webber, "A theory of patterns of passby noise," J. Sound Vib. **262**, 1047–1056 (2003).
[13]J. Beaumont, B. Robin, and C. Semidor, "Event descriptors for qualifying the urban sound environment," in Proceedings of the Forum Acusticum, Budapest, Hungary (2005).
[14]B. De Coensel and D. Botteldooren, "The quiet rural soundscape and how to characterize it," Acta. Acust. Acust. **92**, 887–897 (2006).
[15]J. Beaumont, "Noise events," in Proceedings of the ICA, Madrid, Spain (2007).
[16]J. C. Baird, K. Harder, and A. Preis, "Annoyance and community noise: Psychophysical model of dose-response relationships," J. Environ. Psychol. **17**, 333–343 (1997).
[17]S. Fidell, S. Teffeteller, R. Horonjeff, and D. M. Green, "Predicting annoyance from detectability of low-level sounds," J. Acoust. Soc. Am. **66**, 1427–1434 (1979).
[18]T. Gjestland and G. Oftedal, "Assessment of noise annoyance: The introduction of a threshold level in L_{eq} calculations," J. Sound Vib. **69**, 603–610 (1980).
[19]A. Kjellberg, U. Landström, M. Tesarz, L. Söderberg, and E. Åkerlund, "The effects of nonphysical noise characteristics, ongoing task and noise sensitivity on annoyance and distraction due to noise at work," J. Environ. Psychol. **16**, 123–136 (1996).
[20]P. D. Schomer and L. R. Wagner, "On the contribution of noticeability of environmental sounds to noise annoyance," Noise Control Eng. J. **44**, 294–305 (1996).
[21]M. Sneddon, K. Pearsons, and S. Fidell, "Laboratory study of the noticeability and annoyance of low signal-to-noise ratio sounds," Noise Control Eng. J. **51**, 300–305 (2003).
[22]J. B. Fritz, M. Elhilali, S. V. David, and S. A. Shamma, "Auditory attention—Focusing the searchlight on sound," Curr. Opin. Neurobiol. **17**, 437–455 (2007).
[23]S. Namba and S. Kuwano, "Measurement of habituation to noise using the method of continuous judgment by category," J. Sound Vib. **127**, 507–511 (1988).
[24]*Habituation, Sensitization, and Behavior*, edited by H. V. S. Peeke and L. Petrinovich (Academic, London, UK, 1984).
[25]D. Schreckenberg, R. Guski, I. Schmaus, U. Moehler, and R. Schuemer, "Annoyance and disturbances due to traffic noise at different times of day," in Proceedings of the Internoise, Prague, Czech Republic (2004).
[26]C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*, 2nd ed. (Wiley, Chichester, West Sussex, UK, 2004).
[27]B. Nölting, *Protein Folding Kinetics: Biophysical Methods*, 2nd ed. (Springer-Verlag, Heidelberg, Germany, 2005).
[28]E. T. Rolls and G. Deco, *Computational Neuroscience of Vision* (Oxford University Press, New York, 2002).
[29]B. De Coensel, D. Botteldooren, B. Berglund, M. E. Nilsson, T. De Muer, and P. Lercher, "Experimental investigation of noise annoyance caused by high-speed trains," Acta. Acust. Acust. **93**, 589–601 (2007).
[30]T. De Muer, D. Botteldooren, B. De Coensel, B. Berglund, M. E. Nilsson, and P. Lercher, "A model for noise annoyance based on notice-events," in Proceedings of the Internoise, Rio de Janeiro, Brazil (2005).
[31]T. De Muer, D. Botteldooren, and P. Lercher, "Event based noise annoyance modeling," in Proceedings of the Forum Acusticum, Budapest, Hungary (2005).
[32]B. De Coensel and D. Botteldooren, "Models for soundscape perception and their use in planning," in Proceedings of the Internoise, Istanbul, Turkey (2007).
[33]D. Botteldooren, L. Dekoninck, B. de Greve, B. De Coensel, and P. Lercher, "Annoyance by combined exposure to noise from road traffic and rail traffic discussed in the framework of the noticing model," in Proceedings of the ICA, Madrid, Spain (2007).
[34]B. Berglund and M. E. Nilsson, "Identification of sounds from traffic," Percept. Mot. Skills **97**, 675–688 (2003).
[35]A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (The MIT Press, Cambridge, MA, 1994).
[36]G. Iverson and R. D. Luce, in *Measurement, Judgment and Decision Making*, edited by M. H. Birnbaum (Academic, New York, 1998), pp. 1–79.
[37]T. Gjestland, "Assessment of annoyance from road traffic noise," J. Sound Vib. **112**, 369–375 (1987).
[38]H. Fastl, "Environmental noise assessment: Psychophysical background," in Proceedings of the Internoise, Dearborn, MI (2002).
[39]M. G. Heinz, H. S. Colburn, and L. H. Carney, "Evaluating auditory performance limits," Neural Comput. **13**, 2273–2338 (2001).
[40]B. Wark, B. N. Lundstrom, and A. Fairhall, "Sensory adaptation," Curr. Opin. Neurobiol. **17**, 423–429 (2007).
[41]J. E. R. Staddon, "On rate-sensitive habituation," Adapt. Behav. **1**, 421–436 (1993).

[42]E. Wagner, M. Florentine, S. Buus, and J. McCormack, "Spectral loudness summation and simple reaction time," J. Acoust. Soc. Am. **116**, 1681–1686 (2004).
[43]E. Pöppel, "A hierarchical model of temporal perception," Trends Cogn. Sci. **1**, 56–61 (1997).
[44]E. Szelag, M. Kanabus, I. Kolodziejczyk, J. Kowalska, and J. Szuchnik, "Individual differences in temporal information processing in humans," Acta Neurobiol. Exp. (Warsz.) **64**, 349–366 (2004).
[45]B. De Coensel, T. De Muer, I. Yperman, and D. Botteldooren, "The influence of traffic flow dynamics on urban soundscapes," Appl. Acoust. **66**, 175–194 (2005).
[46]B. De Coensel, D. Botteldooren, F. Vanhove, and S. Logghe, "Microsimulation based corrections on the road traffic noise emission near intersections," Acta. Acust. Acust. **93**, 241–252 (2007).
[47]A. D. May, *Traffic Flow Fundamentals* (Prentice-Hall, Upper Saddle River, NJ, 1990).
[48]D. Helbing, "Fundamentals of traffic flow," Phys. Rev. E **55**, 3735–3738 (1997).
[49]R. F. Voss and J. Clarke, "'$1/f$ noise' in music and speech," Nature (London) **258**, 317–318 (1975).
[50]B. De Coensel, D. Botteldooren, and T. De Muer, "$1/f$ noise in rural and urban soundscapes," Acta. Acust. Acust. **89**, 287–295 (2003).
[51]D. Botteldooren, B. De Coensel, and T. De Muer, "The temporal structure of urban soundscapes," J. Sound Vib. **292**, 105–123 (2006).
[52]H. G. Jonasson, "Acoustical source modelling of road vehicles," Acta. Acust. Acust. **93**, 173–184 (2007).
[53]"Calculation methods for noise immission caused by industry and railway traffic (BeGIS)," Laboratory of Acoustics and Thermal Physics Technical Report, K. U. Leuven, Leuven, Belgium, 1999.
[54]J. M. Fields and J. G. Walker, "Comparing the relationships between noise level and annoyance in different surveys: A railway vs. aircraft and road traffic comparison," J. Sound Vib. **81**, 51–80 (1982).
[55]U. Moehler, "Community response to railway noise: A review of social surveys," J. Sound Vib. **120**, 321–332 (1988).
[56]T. Morihara, T. Sato, and T. Yano, "Comparison of dose-response relationships between railway and road traffic noises: The moderating effect of distance," J. Sound Vib. **277**, 559–565 (2004).
[57]P. Lercher, D. Botteldooren, B. de Greve, L. Dekoninck, and J. Rüdisser, "The effects of noise from combined traffic sources on annoyance: The case of interactions between rail and road noise," in Proceedings of the Internoise, Istanbul, Turkey (2007).
[58]J. M. Fields, "Effect of personal and situational variables on noise annoyance in residential areas," J. Acoust. Soc. Am. **93**, 2753–2763 (1993).
[59]R. Guski, "Personal and social variables as co-determinants of noise annoyance," Noise Health **1**, 45–56 (1999).
[60]P. Lercher, "Environmental noise: A contextual public health perspective," in *Noise and Its Effects*, edited by L. M. Luxon and D. Prasher (Wiley, London, UK, 2007), pp. 345–377.

Regularization in global sound equalization based on effort variation

Nick Stefanakis[a] and John Sarris
School of Electrical and Computer Engineering, National Technical University of Athens, Heroon Polytechniou 9, 157 73 Athens, Greece

Finn Jacobsen
Department of Electrical Engineering, Acoustic Technology, Technical University of Denmark, Ørsteds Plads, Building 352, DK-2800 Kongens Lyngby, Denmark

(Received 20 June 2008; revised 14 May 2009; accepted 28 May 2009)

Sound equalization in closed spaces can be significantly improved by generating propagating waves that are naturally associated with the geometry, as, for example, plane waves in rectangular enclosures. This paper presents a control approach termed effort variation regularization based on this idea. Effort variation equalization involves modifying the conventional cost function in sound equalization, which is based on minimizing least-squares reproduction errors, by adding a term that is proportional to the squared deviations between complex source strengths, calculated independently for the sources at each of the two walls perpendicular to the direction of propagation. Simulation results in a two-dimensional room of irregular shape and in a rectangular room with sources randomly distributed on two opposite walls demonstrate that the proposed technique leads to smaller global reproduction errors and better equalization performance at listening positions outside of the control region compared to effort regularization and compared to a simple technique that involves driving groups of sources identically.
 [DOI: 10.1121/1.3158926]

PACS number(s): 43.55.Br, 43.38.Md, 43.60.Pt [NX] Pages: 666–675

I. INTRODUCTION

The purpose of equalization in room acoustics is to compensate for the undesired modification that an enclosure introduces to signals as, for example, audio or speech. Traditional multi-channel methods introduce digital filters to preprocess the input signal before it is fed to a set of loudspeakers so that the spectral coloration and the reverberation tail associated with the transmission path are reduced.[1,2] Generally, equalization is focused on improving two different attributes of the listening response: the reverberation time and the magnitude response. Modal equalization attempts to control the modal decay of low-frequency modes so that they correspond to a target reverberation time.[3,4] This method is based on a rearrangement of the poles of the listening response. It is a spatially robust method in the sense that increasing the decay rate of a modal resonance at a single position in the room results in an increment of the decay rate at other positions as well.[5] Magnitude response equalization attempts to reduce the unevenness associated with the peaks and dips in the spectrum of the listening response. Generally, the process aims to design the source input signal so that the obtainable signals at a set of receiving positions approximate a set of desired signals.[6,7]

Based on the latter approach, it has been shown that equalization in a large region of a room can be favored by the reproduction of a freely propagating plane wave.[8–13] This is achieved by appropriate positioning of sound sources at two opposite walls perpendicular to the direction of propagation. In this way the equalization can be extended to a spatial region that covers almost the complete volume of the room. It has been shown theoretically that for the successful generation of a plane wave in a rectangular room, the sources at each side are in phase and their amplitudes assume well-defined ratios.[10] These ratios reflect the natural effect of the nearby reflecting surfaces, varying according to the source position inside the room. Also, for the case of a perfectly rectangular room with uniform acoustic wall admittance, optimum source locations as well as favorable listening planes have been identified.[13] In the same work it has been shown that by exploiting the symmetries in the rectangular room, an ideal source distribution can be found. In that case, a single equalization filter at each side and a limited control sensor array are enough to achieve a successful equalization over an area covering almost the complete volume of the room. These observations suggest that the source filters at each one of the two "playing" walls are highly correlated and this correlation is desired in order to achieve a spatially extended equalization.

This work presents a modification to the cost function used in the traditional least-squares approximation. The proposed technique, termed *effort variation regularization*, is based on the general form of Tikhonov regularization[14] and works by penalizing the squared deviations between source strengths.

In the frequency domain, penalization of the complex source strength variations can be realized using square tridi-

[a]Author to whom correspondence should be addressed. Electronic mail: nstefan@mobile.ntua.gr

agonal or bidiagonal differential operators instead of the unitary matrix used in traditional effort regularization.[14] Simulation results for a room of irregular shape demonstrate the applicability of the technique when the conditions for equalization are more difficult than in perfectly rectangular rooms.

Following the traditional least-squares approach for the equalization of broadband signals, effort variation regularization can easily be applied in the time domain using finite impulse response (FIR) equalization filters. Simulation results for a three-dimensional rectangular room show that good global equalization can be achieved with a few control sensors placed outside the listening area, without restraining the presence and motion of listeners inside the room.

II. CONTROL MODEL

Suppose that it is desired to equalize the sound field in a spatial region in an enclosure with L reproduction sources. The sound pressure in the region is sampled by M monitor sensors placed at $\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_M\}$, and this provides a measure of the performance of reproduction in the entire listening space. The pressure at the monitoring sensors subject to the L source excitations can be written as[7]

$$\mathbf{p}_M = \mathbf{Z}_M \mathbf{q}, \tag{1}$$

where $\mathbf{p}_M$ is a column vector with the M complex sound pressures at the monitor sensors (Pa), $\mathbf{q}$ is a column vector with the complex strengths of the L sources (m^3/s), and $\mathbf{Z}_M$ is an $M \times L$ transfer matrix with the transfer functions between the L sources and the M monitor sensors. A small group of N control sensors is selected from the M monitor sensors at $\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\}$, covering a small region centered inside the listening area. It is assumed that this compact control sensor array represents a more practical sound reproduction system that occupies less space and requires less equipment and input channels. This system is informed about the performance of reproduction in the controlled region by the difference between the desired sound pressure and the actual reproduced sound pressure at the control sensors as expressed by

$$\mathbf{e} = \mathbf{p}_d - \mathbf{Z}\mathbf{q}, \tag{2}$$

where $\mathbf{p}_d$ is the vector with the desired sound pressures at the N control sensors, and $\mathbf{Z}$ is the transfer matrix with the transfer functions between the L sources and the N control sensors.

A. Regularization techniques

Equalization is related to the inverse problem of reconstructing the strengths of a number of sources given the transfer matrix and the desired responses at a number of receiving positions in the room. When the number of sensors is less than the number of sources the linear problem is underdetermined and reconstruction of the source strengths requires the use of a regularization technique. One of the most common techniques used in active control is effort regularization, which is based on the standard form of Tikhonov regularization[14–16] and the cost function

$$J^{(\mu)} = (\mathbf{Z}\mathbf{q} - \mathbf{p}_d)^H (\mathbf{Z}\mathbf{q} - \mathbf{p}_d) + \mu \mathbf{q}^H \mathbf{q}, \tag{3}$$

where μ is a positive scalar that weights the penalty term in the cost function. Minimization of this cost function gives a solution to the underdetermined system. Furthermore, it has been shown that a proper choice of the regularization parameter μ has the ability to enlarge the effective area of equalization inside a room.[12] The optimum source strength vector for this regularization technique is

$$\mathbf{q}_o^{(\mu)} = (\mu \mathbf{I} + \mathbf{Z}^H \mathbf{Z})^{-1} \mathbf{Z}^H \mathbf{p}_d. \tag{4}$$

In search of more efficient penalization techniques, power output regularization has recently been presented and associated with the general form of Tikhonov regularization.[12] Used in a similar equalization task, penalization of the total power output favored the reproduction of the plane propagating wave in the room, resulting in an increment of the effective equalization area.

Based also on the general form of Tikhonov regularization, effort variation regularization relies on the use of the cost function

$$J^{(h)} = (\mathbf{Z}\mathbf{q} - \mathbf{p}_d)^H (\mathbf{Z}\mathbf{q} - \mathbf{p}_d) + h(\mathbf{D}\mathbf{q})^H (\mathbf{D}\mathbf{q}), \tag{5}$$

where h is a positive scalar that weights the effort variation penalty term in the cost function, and the weight matrix $\mathbf{D}$ represents a discrete first or second order differential operator. Such matrices are usually of the form[17]

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & -1 & 1 & \cdots \\ 0 & 0 & 0 & -1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{6}$$

and

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots \\ 1 & -2 & 1 & 0 & \cdots \\ 0 & 1 & -2 & 1 & \cdots \\ 0 & 0 & 1 & -2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{7}$$

for first and second order variation penalties, respectively. If the quantity $(\mathbf{D}\mathbf{q})^H(\mathbf{D}\mathbf{q})$ is expanded in terms of the elements of the vector $\mathbf{q}$, it can be seen that such matrices penalize variations between the strengths of adjacent sources and thus force the system to a more uniform solution, which in the extreme case would lead to all the sources having equal strengths ($q_1 = q_2 = \cdots = q_L$). Under the condition that the matrix $h\mathbf{D}^H\mathbf{D} + \mathbf{Z}^H\mathbf{Z}$ is invertible, the optimum solution can be derived as

$$\mathbf{q}_o^{(h)} = (h\mathbf{D}^H\mathbf{D} + \mathbf{Z}^H\mathbf{Z})^{-1}\mathbf{Z}^H\mathbf{p}_d. \tag{8}$$

It should be noted that for such differential operators the matrix $\mathbf{D}^H\mathbf{D}$ is singular, and therefore too strong penalization can be expected to harm the calculation of $(h\mathbf{D}^H\mathbf{D} + \mathbf{Z}^H\mathbf{Z})^{-1}$. Also, in contrast to the case of effort regularization, the numbering of the sources and their positioning in the source strength vector $\mathbf{q}$ is of great importance, and, as will be

shown, it must be consistent with the source array geometry.

B. Reproduction error criteria

The achieved quality of the equalization is measured in the entire listening space with the use of the M monitor sensors. The global reproduction error is calculated as the square root of the mean square value of the errors at the monitor sensors normalized by the energy of the desired sound field at the monitor sensors $\mathbf{p}_{d,M}$,

$$\mathrm{ELS}_m^{(j)} = \left(\frac{\mathbf{e}_M^H \mathbf{e}_M}{\mathbf{p}_{d,M}^H \mathbf{p}_{d,M}} \right)^{1/2} = \left(\frac{(\mathbf{p}_{d,M} - \mathbf{Z}_M \mathbf{q}_o^{(j)})^H (\mathbf{p}_{d,M} - \mathbf{Z}_M \mathbf{q}_o^{(j)})}{\mathbf{p}_{d,M}^H \mathbf{p}_{d,M}} \right)^{1/2}. \quad (9)$$

Here $\mathbf{e}_M = \mathbf{p}_{d,M} - \mathbf{Z}_M \mathbf{q}_o^{(j)}$ and $j = \mu$ or h, which implies effort regularization and effort variation regularization, respectively. For the given definition of the reproduction error, and assuming that the equalization is achieved by the generation of a propagating plane wave, a value up to 0.3 implies that the deviations between the reproduced sound pressure and the desired pressure are within an interval of ± 3 dB, whereas a value up to 0.5 implies deviations within ± 6 dB.

The performance of the reproduction is also measured at the N control sensors as

$$\mathrm{ELS}_c^{(j)} = \left(\frac{(\mathbf{p}_d - \mathbf{Z}\mathbf{q}_o^{(j)})^H (\mathbf{p}_d - \mathbf{Z}\mathbf{q}_o^{(j)})}{\mathbf{p}_d^H \mathbf{p}_d} \right)^{1/2}. \quad (10)$$

In addition to the control approaches described above it is also interesting to define an *ideal system* that uses the information from all M monitor sensors in the adaptation of the optimum source strengths by minimizing the global cost function

$$J^{(\mathrm{id})} = \mathbf{e}_M^H \mathbf{e}_M. \quad (11)$$

The optimum source strengths for this control approach are given by

$$\mathbf{q}_o^{(\mathrm{id})} = (\mathbf{Z}_M^H \mathbf{Z}_M)^{-1} \mathbf{Z}_M^H \mathbf{p}_{d,M}. \quad (12)$$

This corresponds to the best that can be achieved in the ideal, unrealistic case where control sensors covering the entire listening space are used. No kind of regularization is included in the calculation of the optimum source strengths given by Eq. (12). Under the examined conditions, the matrix $\mathbf{Z}_M$ is always overdetermined and $\mathbf{Z}_M^H \mathbf{Z}_M$ is positive definite. The global reproduction performance of the ideal system can also be judged by the global reproduction error by substituting $\mathbf{q}_o^{(j)}$ with $\mathbf{q}_o^{(\mathrm{id})}$ in Eq. (9).

III. SIMULATIONS IN A TWO-DIMENSIONAL ROOM

A. Conditions for the simulations

The purpose of this section is to investigate equalization in a non-rectangular room, which is modeled using the boundary element method.[18,19] A sum of 3784 linear triangular elements is used in order to mesh the physical boundary of the room, shown in Fig. 1. The vertical dimension of the room is kept relatively small ($L_z = 0.1$ m) so that the equalization problem is actually reduced to a two-dimensional one. A constant acoustic wall impedance of $120\rho_0 c$, where $\rho_0 = 1.204$ kg/m^3 is the density of air and $c = 344$ m/s is the speed of sound, has been assumed. Eleven sound sources with monopole characteristics are used to control the sound field; five point sources are placed on the left wall at $x_L = 0.1$ m, and the remaining six sources are placed near the right wall at $x_R = 3.4$ m. The equalization is examined in the listening area, which is defined as a rectangle with the lower left corner at (0.7, 0.3, 0.1) m and the right upper corner at (2.8, 1.8, 0.1) m. A sum of 352 monitor sensors is spread inside the listening area in order to monitor the sound pressure, and a square grid of nine control sensors is used near the middle of the room in order to optimize the complex source strengths. The central control sensor is placed at (2, 1, 0.1) m. The distance between the monitor sensors as well as between the control sensors is 0.1 m in both the x- and the y-direction. In this configuration the control region covers less than 4% of the listening area. The purpose of the arrangement is to reproduce a plane wave that travels in the x-direction, simulating a sound field of constant amplitude that varies as e^{-jkx}.

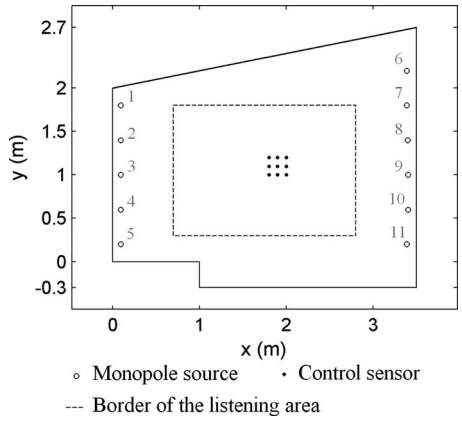


FIG. 1. Configuration of the equalization system in a two-dimensional non-rectangular room. The lower left monopole source coordinates are (0.1, 0.2, 0.1), and the upper left are (0.1, 1.8, 0.1), while the lower right and upper right source coordinates are (3.4, 0.2 0.1) and (3.4, 2.6, 0.1), respectively. The source spacing is 0.4 m at both sides.

B. The ideal system

Reproduction of the plane wave subject to minimization of the reproduction error at the M monitor sensors is accomplished successfully. Implementation of the ideal system is based on knowledge of the transfer functions from all sources to all monitor sensors, something that would be impractical in a real situation. However, this approach is worth examining because it provides useful observations about the correlation in the optimum complex source strengths required for the generation of the plane wave. The optimum source amplitudes are shown in Figs. 2(a) and 2(c) for the five left and the six right point sources, respectively, and the optimum phases are shown in Figs. 2(b) and 2(d). Some deviations between the source amplitudes can be seen in Figs. 2(a) and 2(c). The deviations are seen to be smaller at the left wall than at the right one. On the other hand, inspec-

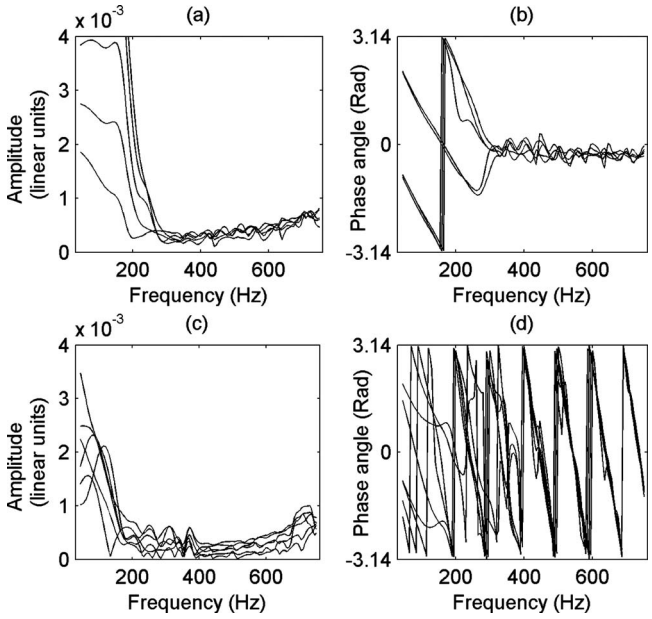


FIG. 2. Amplitudes (a) and phases (b) of the left sources (at $x=0.1$ m), and amplitudes (c) and phases (d) of the right sources (at $y=3.4$ m) with the ideal system.

tion of Figs. 2(b) and 2(d) shows that the sources at each side tend to have identical phases, in particular, above 300 Hz. Note that the phases at the left side are stabilized around zero whereas the phases at the right side tend to vary linearly with the frequency according to $-e^{-jkx_R}$. This means that the sources are 180° out of phase with the desired sound pressure at $x=x_R$, which is necessary to avoid reflections from the receiving wall. This behavior, which has already been observed and explained theoretically by Santillán[8] for the case of rectangular rooms, remains a desired condition for good global equalization also in irregularly shaped rooms.

C. Implementation of effort variation regularization

For the given arrangement and the source numbering shown in Fig. 1, the effort variation penalty term of Eq. (5) is implemented with the matrix

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_L & \mathbf{0}_{6\times 6} \\ \mathbf{0}_{5\times 5} & \mathbf{D}_R \end{bmatrix}, \tag{13}$$

where

$$\mathbf{D}_L = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \tag{14}$$

$$\mathbf{D}_R = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \tag{15}$$

and $\mathbf{0}_{5\times 5}$ and $\mathbf{0}_{6\times 6}$ are zero matrices. These second order differential operators penalize the variation between the strengths of adjacent sources. Note that with the numbering shown in Fig. 1 sources 1 and 5 have only one neighbor and are therefore linked only to sources 2 and 4, respectively, whereas source 2, for example, is linked to both sources 1 and 3. It is also interesting to observe from Eq. (13) that none of the sources 1–5 is linked to any of the opposite sources 6–11. This suggests that the deviations in the solution are penalized independently at each side.

D. Equalization with two coupled source arrays

Inspired by previous work on sound equalization in a rectangular room[13] the idea of coupling the sources at each side of the room is also examined. In this particular case this means that the five left sources are driven with the same amplitude and phase, and the same holds for the six right sources. This technique is evidently similar to using two independent "column" loudspeakers,[20] one at each side of the room. In the frequency domain, only two complex source strengths, one for the left source array and one for the right one, should be estimated. This corresponds to solving an overdetermined system

$$\mathbf{q}_{\text{coupled}} = \begin{bmatrix} q_L \\ q_R \end{bmatrix} = (\mathbf{Z}^H_{\text{coupled}}\mathbf{Z}_{\text{coupled}})^{-1}\mathbf{Z}^H_{\text{coupled}}\mathbf{p}_d, \tag{16}$$

where q_L and q_R are the left and right coupled source array strengths, respectively, and $\mathbf{Z}_{\text{coupled}}$ is the 9×2 matrix carrying the acoustic transfer functions from each source array to the control sensors inside the room. This approach thus assumes a-priori known conditions about the source strengths, giving results that would be similar to those derived by effort variation regularization in the extreme case where $h\rightarrow\infty$.

E. Reproduction performance

The system is now tested for global equalization subject to minimization of reproduction errors only at the nine control sensors. Examination of the reproduction error at the control sensor locations shows that the increment of the regularization parameters μ and h is followed by reduction in the quality of equalization at these positions. Nevertheless, the accuracy of the reproduction in the entire listening area is improved for a non-trivial value of the regularization parameters with both the proposed and the conventional technique. This can be seen in Fig. 3, where the global reproduction error is plotted as a function of the regularization parameter for effort regularization and effort variation regularization at 520 Hz. The values of the regularization parameters in Eqs. (4) and (8) were set equal to $\mu=450$ and $h=900$, and the global reproduction error was determined between 30 and

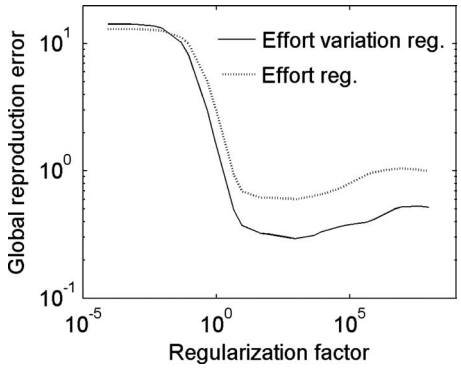


FIG. 3. Global reproduction error as a function of the regularization parameter for effort regularization and effort variation regularization for second order differential operator at 520 Hz.

750 Hz. Figure 4(a) shows the global reproduction error also for the ideal and for the coupled source system, and Fig. 4(b) shows the reproduction error at the control sensor locations for both regularization techniques as well as for the coupled source system. Both regularization techniques achieve an almost perfect equalization result in the control region in the entire frequency range. However, whereas the reproduction error in the control region is of the same order of size both with effort regularization and effort variation regularization, the global performance of these two techniques is evidently different. It can be seen in Fig. 4(a) that effort regularization leads to large global reproduction errors above 300 Hz. This is translated into a rapid decline of the reproduction performance when moving away from the control positions. Whereas this deterioration increases with the frequency with the traditional regularization technique it can be seen that the proposed regularization technique achieves much smaller reproduction errors from 300 to 700 Hz. This indicates that the proposed control approach gives better equalization results at the monitor positions outside the control region. This demonstrates an important advantage in this frequency region, but it should be mentioned that all techniques apart from the ideal system exhibit a poor global performance around 230 Hz. The frequency of this error peak appears to be depending on the location of the control sensors. In any case it was observed that proper adjustment of the regularization factors μ and h reduces the global reproduction error below 0.5 around 230 Hz both for effort regularization and effort variation regularization. In the simulation results presented above, a constant value of the regularization factor was used in the entire frequency range.

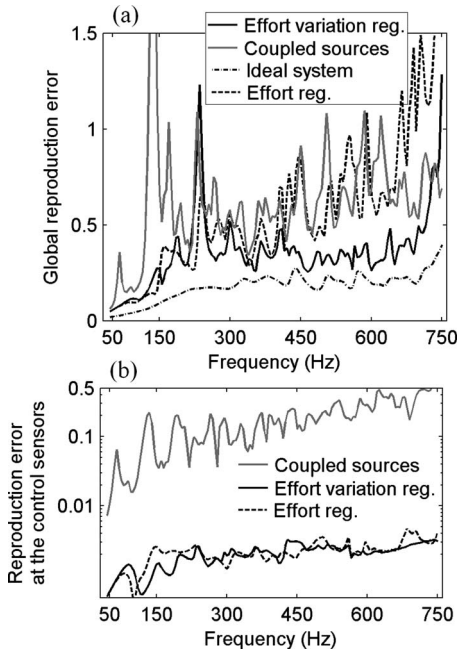


FIG. 4. (a) Global reproduction error as a function of the frequency for the ideal system, effort regularization, effort variation regularization, and coupled source arrays. (b) Reproduction error at the control sensors for effort regularization, effort variation regularization, and coupled source array.

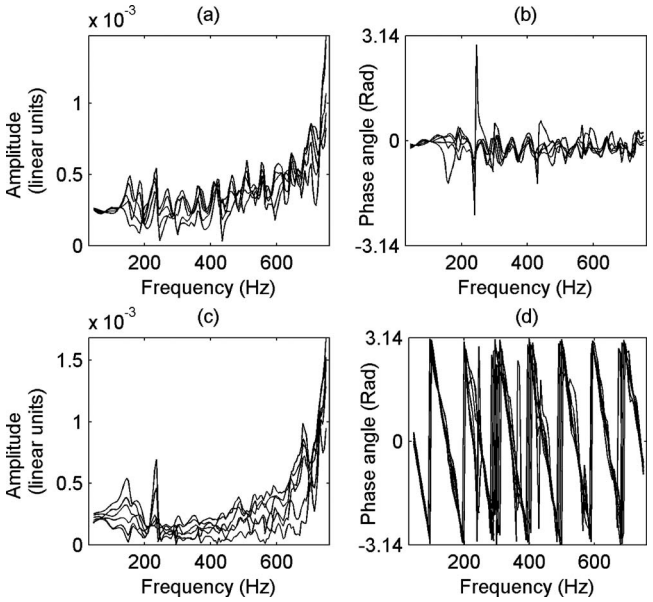


FIG. 5. Amplitudes (a) and phases (b) of the left sources, and amplitudes (c) and phases (d) of the right sources with effort variation regularization.

Examination of the coupled source system shows that it is unable to provide a good global equalization result, although some improvement is observed at the highest frequencies of the investigation. This indicates that the success of the proposed technique compared to the case of the coupled source array is due to imposing a desired correlation without necessarily preventing deviations that are required for the adaptation to the particulars of the room. The individual effect of each regularization technique in each derived solution $\mathbf{q}_o^{(h)}$ and $\mathbf{q}_o^{(\mu)}$ can be seen in Figs. 5 and 6. The variation in the source amplitudes at the left and the right side can be seen in (a) and (b), and the variation in the source phases is shown in (c) and (d). Comparing the source phases shown in Figs. 5 and 6 shows that the effort variation penalty term has reduced the phase deviations at each side. However, whereas the amplitude deviations at the left side are reduced compared to effort regularization, an evident deviation is observed at the right source amplitudes. In agreement with previous observations this shows that effort variation has acted by equalizing the source phases without preventing the amplitude deviation that is necessary for room compensation.

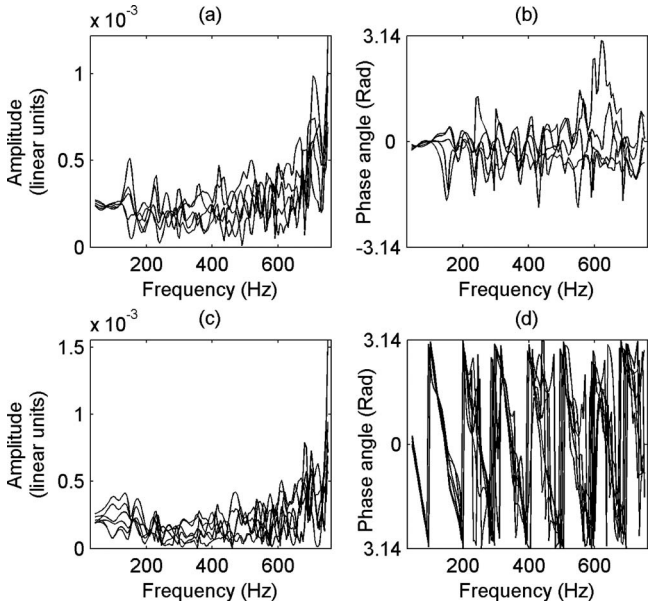


FIG. 6. Amplitudes (a) and phases (b) of the left sources, and amplitudes (c) and phases (d) of the right sources with effort regularization.

IV. EFFORT VARIATION REGULARIZATION IN TIME DOMAIN EQUALIZATION OF BROADBAND SIGNALS

The following is intended to demonstrate the use of effort variation regularization in the time domain and also involves applying the proposed technique to cases of more complicated source distributions, such as those required for three-dimensional rooms. To overcome the time and memory restrictions imposed by the boundary element method, the acoustic transfer functions are calculated here on the basis of an analytical model which is appropriate for rectangular rooms.

A. Time domain implementation of equalization

The following analysis is based on the matrix formulation by Elliott and Nelson.[6] A block diagram of the equalization system is shown in Fig. 7. The aim is to design L digital FIR control filters, one for each sound source, with impulse responses $h_l(n)$ such that the obtainable signal $\hat{d}_m(n)$ at the mth control sensor $(1 \le m \le N)$ is the best approximation to the desired signal $d_m(n)$ at the sensor. (Although the index m has so far been associated with the monitor sensors, it is used here to avoid confusion with the discrete time index n.) Here $d_m(n)$ is a delayed version (by δ_m samples) of the original

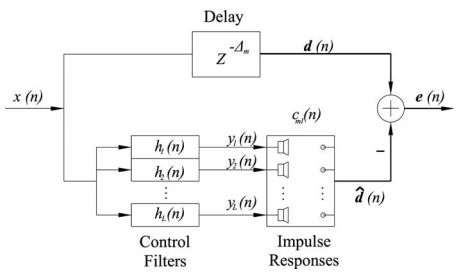


FIG. 7. Block diagram for equalization of broadband signals.

input signal $x(n)$. Since a plane wave traveling in the y-direction is desired, the duration of the delay depends on the ordinate of the listening position corresponding to the propagation of the plane wave.

The impulse response between the input of the lth source and the output of the mth sensor is modeled by an FIR filter with J coefficients, represented by $c_{ml}(n)$ in Fig. 7. The signal $\hat{d}_m(n)$ detected by the mth sensor can be expressed as

$$\hat{d}_m(n) = \sum_{l=1}^{L} \sum_{i=1}^{I} h_l(i) r_{ml}(n-i), \tag{17}$$

where $h_l(i)$ is the ith coefficient of the FIR control filter whose output is the input to the lth loudspeaker, and $r_{ml}(n)$ is the reference signal produced by convolving the input signal $x(n)$ with the impulse response $c_{ml}(n)$. Each control filter is assumed to have an impulse response of I samples. In vector notation Eq. (17) can be written as

$$\hat{d}_m(n) = \mathbf{h}^T(0)\mathbf{r}_m(n) + \mathbf{h}^T(1)\mathbf{r}_m(n-1) + \cdots + \mathbf{h}^T(I-1)\mathbf{r}_m(n-I+1), \tag{18}$$

where a composite tap weight vector and a reference signal vector have been defined by

$$\mathbf{h}^T(i) = \begin{bmatrix} h_1(i) & h_2(i) & \cdots & h_L(i) \end{bmatrix} \tag{19}$$

and

$$\mathbf{r}_m^T(n) = \begin{bmatrix} r_{m1}(n) & r_{m2}(n) & \cdots & r_{mL}(n) \end{bmatrix}. \tag{20}$$

The final vector that contains all the I coefficients of all the L control FIR filters can be written as

$$\mathbf{a}^T = \begin{bmatrix} \mathbf{h}^T(0) & \mathbf{h}^T(1) & \cdots & \mathbf{h}^T(I-1) \end{bmatrix}, \tag{21}$$

and the error vector signal at the N control sensors can be expressed as

$$\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{R}(n)\mathbf{a}, \tag{22}$$

where

$$\mathbf{R}(n) = \begin{bmatrix} \mathbf{r}_1^T(n) & \mathbf{r}_1^T(n-1) & \cdots & \mathbf{r}_1^T(n-I+1) \\ \mathbf{r}_2^T(n) & \mathbf{r}_2^T(n-1) & \cdots & \mathbf{r}_2^T(n-I+1) \\ & & \vdots & \\ \mathbf{r}_M^T(n) & \mathbf{r}_M^T(n-1) & & \mathbf{r}_M^T(n-I+1) \end{bmatrix} \tag{23}$$

is the matrix of filtered reference signals. The optimal coefficients of the control filters are determined by minimizing a performance index defined by[21]

$$J = E\{\mathbf{e}^T(n)\mathbf{e}(n) + \lambda \mathbf{a}^T(n)\mathbf{W}\mathbf{a}(n)\}, \tag{24}$$

where E is the expectation operator, $\mathbf{W}$ is a generally symmetric matrix that defines the type of penalization in the cost function, and λ is the regularization parameter that defines the weighting assign to the penalty term. It can be seen that the performance index can be expressed as a quadratic function in terms of all the individual coefficients in the equalization filters. This performance index has a unique global minimum that corresponds to the optimal control filter coefficients

$$\mathbf{a}_o = (E\{\mathbf{R}(n)^T\mathbf{R}(n) + \lambda\mathbf{W}\})^{-1}E\{\mathbf{R}(n)^T\mathbf{d}(n)\}. \quad (25)$$

It should be noticed that Eq. (25) gives the optimal solution in a statistical sense. The only required information about the source input signal is the autocorrelation function. In the case when $\mathbf{W}=\mathbf{I}$ (the identity matrix), the second term in the expression for the cost function, Eq. (24), becomes identical to the effort penalty term that has been used in similar active control applications.[14,15,21] The optimum control filters in this case are specified in terms of the value of the regularization parameter μ as

$$\mathbf{a}_o^{(\mu)} = (E\{\mathbf{R}(n)^T\mathbf{R}(n) + \mu\mathbf{I}\})^{-1}E\{\mathbf{R}(n)^T\mathbf{d}(n)\}. \quad (26)$$

The transformation of the proposed technique to the form necessary for the time domain equalization of random signals is easily accomplished with proper weighting of the filter coefficients. Again the differential operators must be constructed so as to be in agreement with the source numbering and the architecture of the filter vector $\mathbf{a}$. Considering that the sources are numbered continuously, as in the example presented previously in the frequency domain, the matrix $\mathbf{W}$ can be straightforwardly composed by similar differential operators $\mathbf{D}$ as those defined in Eqs. (14) and (15),

$$\mathbf{W} = \mathbf{\Delta}^T\mathbf{\Delta}, \quad (27)$$

where $\mathbf{\Delta}$ is square matrix of order $I \cdot L$ defined as

$$\mathbf{\Delta} = \begin{bmatrix} \mathbf{D} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D} \end{bmatrix}. \quad (28)$$

By inspecting the structure of $\mathbf{\Delta}$ and the architecture of the composite filter vector $\mathbf{a}$ in Eqs. (19) and (21) it can be seen that the penalty term in the cost function in Eq. (24) can be written as

$$\mathbf{a}^T\mathbf{\Delta}^T\mathbf{\Delta}\mathbf{a} = \sum_{i=0}^{I-1} \mathbf{h}^T(i)\mathbf{D}^T\mathbf{D}\mathbf{h}^T(i). \quad (29)$$

This implies that the symmetrical matrix $\mathbf{W}$ used in the time domain measures the deviation between adjacent source filter coefficients in the same way as the matrix $\mathbf{D}^T\mathbf{D}$ measures the deviation between adjacent source strengths in the frequency domain. Under the condition that the matrix $E\{\mathbf{R}(n)^T\mathbf{R}(n) + h\mathbf{\Delta}^T\mathbf{\Delta}\}$ is invertible, the optimum filter coefficients can be computed by substituting Eq. (27) into Eq. (25).

B. Conditions for the simulations

The simulations that follow are intended for demonstrating the benefits of the proposed technique in time domain equalization of random signals in a three-dimensional rectangular room. For this purpose, the conventional modal sum of the sound field in a lightly damped rectangular enclosure with walls of uniform specific acoustic admittance proposed by Morse[22] is used in the form presented by Bullmore *et al.*[23] The sources are modeled as square pistons that vibrate with a normal velocity $u_l = q_l/A$, where $A=a^2$ is the area of the piston sources. The piston sources are oriented inside the room so that their surfaces are parallel to the xz-plane.

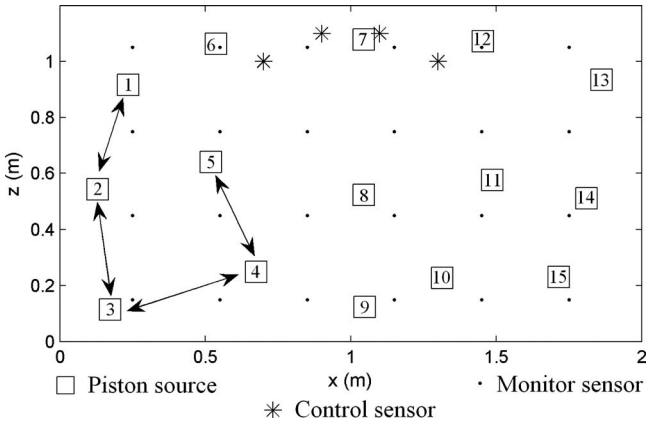


FIG. 8. Front view of the room and of the equalization system. The black arrows indicate the correlation path while moving from source 1 to source 15 on the left wall at $y=0.1$ m.

The room has dimensions of $L_x=2$ m, $L_y=3.2$ m, and $L_z=1.2$ m, and the damping ratio is set equal to 0.015, which corresponds to a reverberation time of 0.73 s at 100 Hz. The sampling rate in the simulations is 1 kHz. Equalization in an extended three-dimensional listening area inside the room is now attempted by the simulation of a plane wave traveling along the y-axis. This is to be achieved with the use of 30 piston sources, 15 on each wall perpendicular to the direction of propagation at $y=0.1$ and $y=3.1$ m. The piston sources are placed according to a 3×5 pattern on each wall, with their centers displaced 9 cm from their initial positions at a random angle on the xz-plane, as shown in Fig. 8. This random source distribution is chosen in order to avoid the ideal conditions that are met by deterministic source placement in the rectangular room. Although proper placement tactics can simplify the problem and define simple optimum correlation in the source equalization filters,[10,13] these conditions are not easily defined in the real world, for example, because of non-ideal room shape, source misplacement, and non-uniform acoustic properties of the boundaries. Any correlation that might be introduced because of the symmetries in the analytical model is here avoided by misplacement of the sources. However, care is taken for the sources not to be close to one another or close to the edges and the corners of the room.

The global reproduction error in these simulations is calculated in a three-dimensional rectangular volume of $1.5\times2.4\times0.9$ m^3. The sound pressure in this volume is sampled by a grid of $6\times8\times4=192$ monitor sensors. The monitor sensors are placed as follows: in eight sensor planes of 6×4 monitor sensors, perpendicular to the direction of propagation, with the first plane placed at $y=0.3$ m and the last one at $y=2.71$ m. Each monitor sensor plane is separated from the other by a distance of 0.343 m, which is equal to the distance that the sound travels in one sampling period. The distance between monitor sensors along the x- and z-axes is equal to 0.3 m and the monitor sensors are extended from $x=0.25$ to $x=1.75$ m and from $z=0.15$ to $z=1.05$ m. The monitor sensors are presented as black dots in Fig. 8.

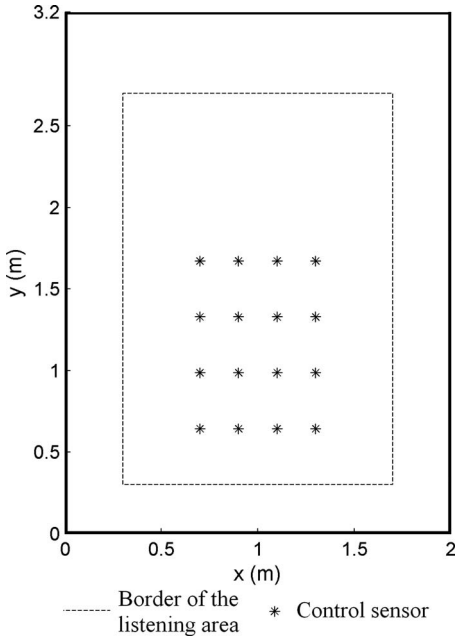


FIG. 9. Top view of the rectangular room.

Although equalization of the sound field is required in an extended three-dimensional region, this is to be achieved with the limited number of $4\times4=16$ control sensors placed, as shown in Figs. 8 and 9. The separation distance between control sensors along the y-axis is also equal to 0.343 m while the separation distance along the x-axis is equal to 0.2 m. The control sensors are located at $x=0.7$, 0.9, 1.1, and 1.3 m with the first four error sensors placed at $y=0.3$ m, and the last four ones at $y=1.329$ m (see Fig. 9). The control sensors at $x=0.7$ and 1.3 m are at a height of $z=1$ m, and those at $x=0.9$ and 1.1 m are at a height of $z=1.1$ m. The control sensors are thus placed just below the ceiling, at a position where in a more realistic case of a higher room they would not interfere with the presence and motion of listeners.

For the time domain simulations, an input signal with a delta function as the autocorrelation function is assumed, and the loudspeakers are modeled as first order analog high pass filters with a pole at 100 Hz. The frequency responses from the input of each loudspeaker to the output of each sensor are calculated with the modal model. All the natural modes up to 1100 Hz are taken into account for the simulation of the sound field in the room. These frequency responses are multiplied by the frequency response of a low pass anti-aliasing filter and by the frequency response of the loudspeakers, and an inverse fast Fourier transform (FFT) is applied for the calculation of the discrete impulse responses $c_{ml}(n)$. The number of coefficients used for each impulse response is $J=300$, and the number of coefficients in each control filter $h_l(n)$ is set equal to $I=70$. A delay of 30 samples is applied to obtain the desired signal at the first four control sensors at $y=0.3$ m, and additional delays of one, two, and three samples are used for the second, third, and fourth lines of control sensors at $y=0.643$, 0.986, and 1.329 m, respectively.

C. Implementation of effort variation regularization

A second order differential operator is constructed that counts the squared deviations of the filter samples for two individual series of 15 sources. A correlation path is specified according to the source numbering, which for the left sources 1–15 at $y=0.1$ m is shown in Fig. 8. It can be seen that sources are linked vertically rather than horizontally. A similar correlation path is applied on the other side, linking sources 16–30 at $y=3.1$ m. A 30×30 differential operator is thus constructed as

$$\mathbf{D}=\begin{bmatrix}\mathbf{D}' & \mathbf{0}\\ \mathbf{0} & \mathbf{D}'\end{bmatrix}, \tag{30}$$

where $\mathbf{D}'$ is a 15×15 tri-diagonal matrix with all elements in the main diagonal equal to -2, except elements (1,1) and (15,15), which are equal to -1, and all elements in the first two parallel diagonals are 1. The optimum equalization filters can now be calculated for effort weighting according to Eq. (26) and for the proposed technique according to Eq. (25), where the matrix $\mathbf{W}$ is constructed according to Eqs. (27) and (28), using $\mathbf{D}$ as in Eq. (30).

D. Reproduction performance

The success of each control technique is judged by the accuracy of reproduction in the entire listening area. The global reproduction error can be calculated as a function of the frequency as

$$\mathrm{ELS}(\omega)=\left(\frac{\sum_{m=1}^{M}(1-|p_m(\omega)|)^2}{M}\right)^{1/2}, \tag{31}$$

where $p_m(\omega)$ is the reproduced sound pressure at the mth monitor sensor at radial frequency ω, and $M=192$ is the number of the monitor sensors. The reproduced sound pressure $p_m(\omega)$ is determined by calculating an FFT of the signal $\hat{d}_m(n)$ detected at each sensor after equalization. This global index is intended for examining the quality of equalization achieved with the limited control sensor array previously presented. Apart from effort regularization and effort variation regularization, equalization with two individual coupled source arrays is also examined. The coupling condition assumes exactly the same equalization filter for all 15 sources at each side, specifying thus an absolute a-priori known condition which is similar to the previous example of equalization in the two-dimensional non-rectangular room. This technique has given excellent results in the case of sound equalization in a three-dimensional room with sources optimally placed on each wall.[13]

The values of the regularization parameters for effort regularization and for effort variation regularization were set equal to $\mu=10^{10}$ and $h=0.5\times10^{12}$, respectively. These values were chosen as the ones that ensure the minimum possible global reproduction error for each technique throughout the entire frequency range from 0 to 500 Hz. The variation in ELS as a function of the frequency for effort regularization, effort variation regularization, coupled source arrays, and ideal system can be seen in Fig. 10.

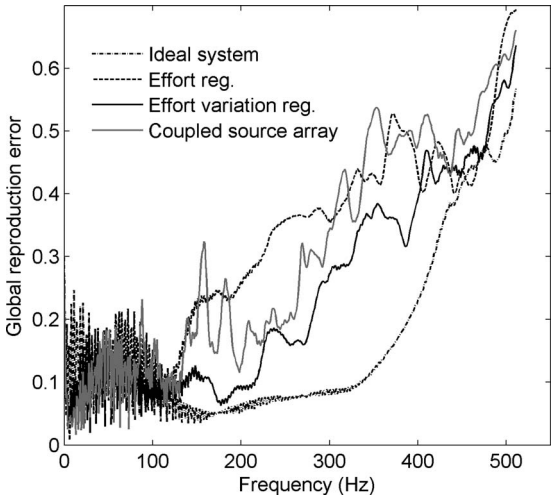


FIG. 10. Global reproduction error as a function of the frequency for effort regularization, effort variation regularization, coupled source arrays, and ideal system calculated in the time domain.

A reduction in the equalization performance with frequency is apparent with all control techniques. The performance of the ideal system indicates that good equalization is possible up to about 500 Hz, where the deviations of the reproduced sound pressure inside the listening area remain below the ±6 dB criterion. Source coupling seems to offer an advantage compared to classical effort regularization, but the global reproduction error of the proposed technique is always smaller than that of effort regularization and also that with the coupled sources. Again, this improvement is translated into reduction in equalization errors at listening positions away from the control sensors. A representative example can be seen in Fig. 11 for the monitor sensor at (1.45, 2.358, 0.45) m, illustrating the impulse and the frequency response before and after equalization for classical effort regularization (a) and (b), coupled source array (c) and (d), and effort variation regularization (e) and (f). The results

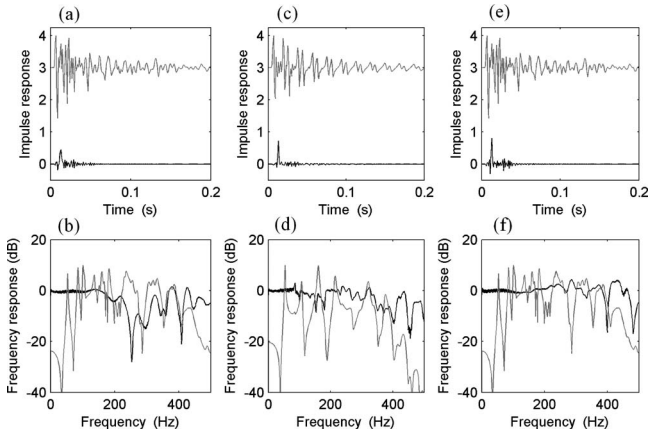


FIG. 11. Impulse and frequency responses at the monitor sensor at (1.45, 2.358, 0.45) m for effort regularization [(a) and (b)], coupled source array [(c) and (d)], and effort variation regularization [(e) and (f)]. The results before equalization in (a), (b), (e), and (f), shown with gray lines, are obtained by exciting the room with only source number 1. The results before equalization in (c) and (d) are obtained after exciting the room with all 15 left sources. The impulse responses before equalization are shown with an offset of +3 linear units for clarity.

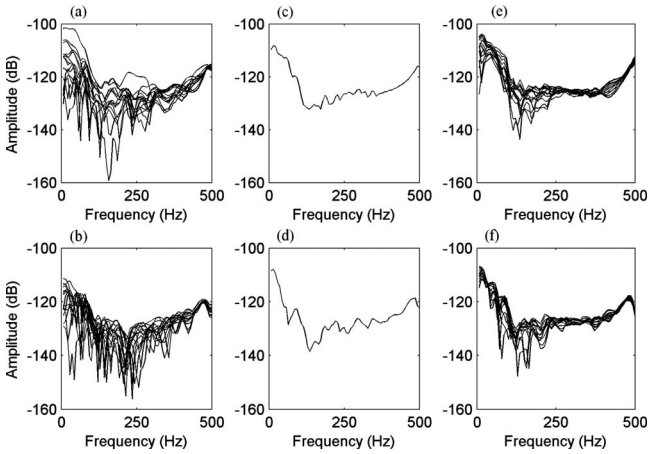


FIG. 12. Source amplitudes as functions of the frequency for effort regularization [(a) and (b)], coupled source array [(c) and (d)], and effort variation regularization [(e) and (f)]. The left sources are shown in (a), (c), and (e), and the right sources are shown in (b), (d), and (f).

before equalization, shown with gray lines, are obtained by exciting the room with piston source number "1" in (a), (b), (e), and (f) and by exciting the room with all left sources at $y=0.1$ m in (c) and (d). The impulse and frequency responses before equalization include the effect of the antialiasing filter. The impulse responses before equalization are shown with an offset of +3 linear units for the sake of clarity.

Some understanding of the effect of each control technique in the estimation of the solution $\mathbf{a}_o$ is gained by observing in Fig. 12 the variation in the amplitude (as derived from the FFT of the equalization filters impulse responses) for effort regularization in (a) and (b), coupled source array in (c) and (d), and effort variation regularization in (e) and (f). The upper row of graphs refers to the left 15 sources, and the lower one to the right sources. Although the proposed technique has reduced the amplitude variation compared to that of the classical approach in (a) and (b), a certain variation can be seen at both sides in (e) and (f) below 250 Hz, which possibly explains the avoidance of the steep error peaks observed for the coupled source case between 100 and 200 Hz in Fig. 10.

V. CONCLUSIONS

A control approach termed effort variation regularization has been proposed and examined. In the frequency domain this technique involves modifying the original cost function based on minimization of the least-squares reproduction error by adding a term that is proportional to the squared deviations between complex source strengths, calculated independently for the sources at the two walls perpendicular to the direction of propagation. Simulation results in a two-dimensional non-rectangular room have shown that effort variation regularization results in equalization of the phase of the complex strengths on each side of the room, without preventing the amplitude deviations that are necessary for room compensation. Applied to the time domain equalization of broadband signals, effort variation regularization has been successfully realized by independently penalizing the squared deviations of the source equalization filter coeffi-

cients at each side of the room. The proposed technique has been demonstrated to lead to smaller global reproduction error and thus better equalization performance at listening positions away from the control region than effort regularization and a simple coupling source array method.

[1]J. Mourjopoulos, P. M. Clarkson, and J. K. Hammond, "A comparative study of least-squares and homomorphic techniques for the inversion of mixed phase signals," in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP' 82), Paris, France (1982), Vol. **7**, pp. 1858–1861.

[2]J. Mourjopoulos, "On the variation and invertibility of room impulse response functions," J. Sound Vib. **102**, 217–228 (1985).

[3]A. Mäkivirta, P. Antsalo, M. Karjalainen, and V. Valimaki, "Modal equalization of loudspeaker-room responses at low frequencies," J. Audio Eng. Soc. **51**, 324–343 (2003).

[4]M. Karjalainen, P. Atsalo, and A. Mäkivirta, "Modal equalization by temporal shaping of room responses," in Proceedings of the AES 23rd International Conference, Copenhagen (2003).

[5]J. Wilson, M. Capp, and R. Stuart, "The loudspeaker-room interface-controlling excitation of room modes," in Proceeding of the AES 23rd International Conference, Copenhagen (2003).

[6]S. J. Elliott and P. A. Nelson, "Multiple-point equalization in a room using adaptive digital filters," J. Audio Eng. Soc. **37**, 899–907 (1989).

[7]F. Asano and D. Swanson, "Sound equalization in enclosures using modal reconstruction," J. Acoust. Soc. Am. **98**, 2062–2069 (1995).

[8]A. O. Santillán, "Spatially extended sound equalization in rectangular rooms," J. Acoust. Soc. Am. **110**, 1989–1997 (2001).

[9]J. C. Sarris, F. Jacobsen, and G. E. Cambourakis, "Sound equalization in a large region of a rectangular enclosure," J. Acoust. Soc. Am. **116**, 3271–3274 (2004).

[10]A. O. Santillán, C. S. Pedersen, and M. Lydolf, "Experimental verification of a low-frequency global sound equalization system based on free field propagation," Appl. Acoust. **68**, 1063–1085 (2007).

[11]A. Celestinos and S. Birkedal Nielsen, "Controlled acoustic bass system (CABS), a method to achieve uniform sound field distribution at low frequencies in rectangular rooms," J. Audio Eng. Soc. **56**, 915–931 (2008).

[12]N. Stefanakis, J. Sarris, G. Cambourakis, and F. Jacobsen, "Power-output regularization in global sound equalization," J. Acoust. Soc. Am. **123**, 33–36 (2008).

[13]N. Stefanakis, J. Sarris, and G. Cambourakis, "Source placement for equalization in small enclosures," J. Audio Eng. Soc. **56**, 357–371 (2008).

[14]P. A. Nelson, "A review of some inverse problems in acoustics," Int. J. Acoust. Vib. **6**, 118–134 (2001).

[15]S. J. Elliott, *Signal Processing for Active Control* (Academic, London, 2001).

[16]P. A. Gauthier, A. Berry, and W. Woszczyk, "Sound-field reproduction in-room using optimal control techniques: Simulations in the frequency domain," J. Acoust. Soc. Am. **117**, 662–678 (2005).

[17]P. C. Hansen, "Perturbation bounds for discrete Tikhonov regularization," Inverse Probl. **5**, L41–L44 (1989).

[18]M. R. Bai, "Study of acoustic resonance in enclosures using eigenanalysis based on boundary element methods," J. Acoust. Soc. Am. **91**, 2529–2538 (1992).

[19]N. Stefanakis and J. Sarris, "Sound field reproduction using the boundary element method," in Proceedings of the ICSV13-Vienna (2006).

[20]D. L. Kepler and D. W. Steele, "Constant directional characteristics from a line source array," J. Audio Eng. Soc. **11**, 198–202 (1963).

[21]P. A. Nelson, F. Orduña-Bustamante, and H. Hamada, "Multichannel signal processing techniques in the reproduction of sound," J. Audio Eng. Soc. **44**, 973–989 (1996).

[22]P. M. Morse, *Vibration and Sound*, 2nd ed. (McGraw-Hill, New York, 1948).

[23]A. J. Bullmore, P. A. Nelson, A. R. D. Curtis, and S. J. Elliott, "The active minimization of harmonic enclosed sound fields, part II: A computer simulation," J. Sound Vib. **117**, 15–33 (1987).

Sound power emitted by a pure-tone source in a reverberation room[a)]

Finn Jacobsen[b)]
Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark, Building 352, DK-2800 Kgs. Lyngby, Denmark

Alfonso Rodríguez Molares
ETSE Telecomunicación, Universidade de Vigo, Campus Lagoas-Marcosende, E-36310 Vigo, Spain

(Received 18 February 2009; revised 5 May 2009; accepted 26 May 2009)

Energy considerations are of enormous practical importance in acoustics. In "energy acoustics," sources of noise are described in terms of the sound power they emit, the underlying assumption being that this property is independent of the particular environment where the sources are placed. However, it is well known that the sound power output of a source emitting a pure tone or a narrow band of noise actually varies significantly with its position in a reverberation room at low frequencies, and even larger variations occur between different rooms. The resulting substantial uncertainty in measurements of sound power as well as in predictions based on knowledge of sound power is one of the fundamental limitations of energy acoustics. The existing theory for this phenomenon is fairly complicated and has only been validated rather indirectly. This paper describes a far simpler theory and demonstrates that it gives predictions in excellent agreement with the established theory. The results are confirmed by experimental results as well as finite element calculations. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3158918]

PACS number(s): 43.55.Cs, 43.58.Bh [AJZ] Pages: 676–684

I. INTRODUCTION

It has been known for many years that sound power emitted by a stationary sound source in a reverberant room depends on the particulars of the room (shape, size, and damping) and the position of the source. The variations are relatively small for sources of broad band noise but fairly substantial for sources that emit pure tones, in particular below the Schroeder frequency. This implies a fundamental contribution to the measurement uncertainty in any sound power measurement that takes place in a non-anechoic room irrespective of whether the diffuse-field method or the sound intensity method is used and, of course, to a similar uncertainty in any prediction of the sound pressure level generated by the source in another room calculated on the basis of the sound power of the source (even if this quantity has been measured under free-field conditions and is very close to the "true" free-field sound power). For pure-tone sources, this fundamental component of the uncertainty is much larger than all other contributions. The only way of reducing it is to determine the sound power at different positions in different rooms or in a room that can be changed, e.g., with a large rotating vane.

II. A BRIEF DESCRIPTION OF THE ESTABLISHED THEORY

The existing theory is essentially due to Lyon[1] and Davy[2] but is later modified in accordance with more recent findings by Weaver.[3] Lyon's approach was based on the analytical Green's function in a rectangular room (and in a point-driven plate), which is a modal sum. Replacing modal sums by integrals and assuming that the modal frequencies have a Poisson distribution (i.e., are distributed independently), he derived the following expression for the normalized standard deviation of the sound power output of a monopole emitting a pure tone in a room,

$$\varepsilon\{P_a\} = \frac{1}{\sqrt{M_s}}\left(\frac{3}{2}\right)^{3/2}, \tag{1}$$

where M_s is the statistical modal overlap,

$$M_s = n(f)B_s = \frac{4\pi V f^2}{c^3}\frac{1}{2\tau} = \frac{12\pi \ln(10) V f^2}{T_{60}c^3} = \frac{S\,\mathrm{Re}\{\beta\}k^2}{\pi} = \frac{Ak^2}{8\pi}, \tag{2}$$

in which $n(f)$ is the modal density (in modes per hertz), B_s is the statistical modal bandwidth (in hertz), V is the volume of the room, f is the frequency, c is the speed of sound, τ is the modal time constant, T_{60} is the corresponding reverberation time, β is the wall admittance normalized by the characteristic impedance of air, S is the surface area of the room, k is the wave number, and A is the total absorption area of the room. (In this expression, one-dimensional (1D) and two-dimensional (2D) modes have been ignored.) Lyon used the

[a)] Portions of this work were presented in "The sound power of sources in reverberant surroundings," Proceedings of NOVEM 2009, Cambridge, England, April 2009.

[b)] Author to whom correspondence should be addressed. Electronic mail: fja@elektro.dtu.dk

 0001-4966/2009/126(2)/676/9/$25.00

equivalent noise modal bandwidth, which is half the statistical bandwidth. Although Lyon's expression at that time was largely interpreted as the source position variance, it is clear that since his statistical method involved different distributions of the modal frequencies the expression is in effect a theoretical estimate of the normalized *ensemble* standard deviation, with variation over rooms—a quantity that is evidently of fundamental importance for the uncertainty but difficult to measure.

Twelve years later, Davy[2] extended Lyon's theory by deriving a more general expression of the power transmission function averaged over multiple source and receiver positions. Assuming a "nearest neighbor" distribution of the modal frequencies, he found a normalized ensemble standard deviation of the output power (or the real part of the radiation impedance) of

$$\varepsilon\{P_a\} = \frac{1}{\sqrt{M_s}}\left(\left(\frac{3}{2}\right)^3 - \frac{1}{2}\right)^{1/2} \tag{3}$$

for high modal overlap, ignoring 1D and 2D modes in a rectangular room. (No simple expression in closed form was derived for the case of low modal overlap.) Some experimental results from measurements of power transmission functions in a very large (600 m^3) room were presented. However, Davy's main issue in this paper was to demonstrate that a very different theory derived by Waterhouse[4] was incorrect, and this he certainly did. The assumption of the modal frequencies having a nearest neighbor distribution rather than being distributed independently came from evidence of a "repulsion" effect between modal frequencies already anticipated by Lyon.[1] The modal frequencies seem to repel each other, and thus their distribution is not completely random but closer to the average density than one might have expected.

Some years later, Davy[5] discussed possible improvements of his theory, but in this paper he was essentially concerned with the variance of random noise passed through a reverberation room and measured using a finite averaging time.

In the late 1980s, Weaver[3] discussed Davy's theory and suggested replacing his K [the term $(3/2)^3$ in Eq. (3)] with 3, "which is appropriate for a Gaussian distribution of amplitudes and based on vague arguments invoking the central limit theorem." At that time, it had been established that modal frequencies tend to exhibit "spectral rigidity" or "long range repulsions" in accordance with the random matrix theory of Gaussian orthogonal ensembles.[6] This seems to be generally accepted now.[7] A value of $K=3$ has also been favored by Lobkis *et al.*[8]

In 1990, Davy[9] modified his theory so as to take account of the Gaussian orthogonal ensemble modal frequency spacing. Equation (3) now became

$$\varepsilon\{P_a\} = \frac{1}{\sqrt{M_s}}\left(\left(\frac{3}{2}\right)^3 - 1\right)^{1/2}. \tag{4}$$

This expression was further discussed in yet another congress paper by Davy[10] from the late 1990s. It can easily be modified so as to take account of 1D and 2D modes. However, it is not completely clear from these papers whether its validity is restricted to high modal overlap, as Eq. (3).

Various consequences of Davy's[2,5,9,10] original and modified theory have been examined experimentally, but there is very little experimental evidence in direct support of Eq. (4).

III. AN ALTERNATIVE THEORY

The alternative theory presented in what follows is far simpler than the theory briefly described in the foregoing and does not even take account of the phenomenon of modes. It was derived independently by Jacobsen[11] and Pierce[12] about 30 years ago.

The sound field in a reverberant room is modeled as an infinite sum of plane waves,

$$p_{\mathrm{rev}}(\mathbf{r}) = \lim_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{n=1}^{N} A_n e^{j(\omega t+\mathbf{k}_n\cdot\mathbf{r})} = \lim_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{n=1}^{N} |A_n| e^{j(\omega t+\varphi_n+\mathbf{k}_n\cdot\mathbf{r})}, \tag{5}$$

where the phase angles φ_n are uniformly distributed between 0 and 2π, the amplitudes $|A_n|$ have an arbitrary distribution, the wave number vectors $\mathbf{k}_n$ are uniformly distributed over all angles of incidence (corresponding to a sinusoidal distribution of the polar angles and a uniform distribution of the azimuth angles), and $\mathbf{r}$ is the observation point. This stochastic pure-tone diffuse-field interference model was originally developed by Waterhouse.[13]

The sound field at the source position ($\mathbf{r}_0$) may be regarded as the sum of the direct field and the reverberant field, and therefore the radiation impedance is the sum of the free-field radiation impedance and the complex ratio of the sound pressure associated with the reverberant field at the source position and the volume velocity of the source,

$$Z_r = \frac{\rho c k^2}{4\pi} + \frac{jk\rho c}{4\pi a} + \frac{p_{\mathrm{rev}}(\mathbf{r}_0)}{Qe^{j\omega t}}, \tag{6}$$

where ρ is the density of air, a is the radius of the monopole that generates the sound field (modeled as a small pulsating sphere), and Q is its volume velocity. (As pointed out by Lyon,[1] the imaginary part of the free-field radiation impedance diverges as the radius of the sphere goes toes to zero.) All phases are equally probable in the reverberant part of the sound field, which leads to the conclusion that on the average the monopole emits its free-field sound power output,

$$E\{P_a\} = \frac{|Q|^2}{2}E\{\mathrm{Re}\{Z_r\}\} = P_{a,\text{free field}} = \frac{\rho c k^2 |Q|^2}{8\pi}. \tag{7}$$

However, because of the third term in Eq. (6) the actual sound power output of the source varies between different outcomes of the stochastic process; i.e., it varies with the source position and it varies from room to room. The corresponding variance can be calculated as follows:

$$\sigma^2\{P_a\} = \frac{|Q|^4}{4}\sigma^2\{\mathrm{Re}\{Z_r\}\} = \frac{|Q|^4}{4}\sigma^2\left\{\mathrm{Re}\left\{\frac{p_{\mathrm{rev}}(\mathbf{r}_0)}{Qe^{j\omega t}}\right\}\right\}$$
$$= \frac{|Q|^2}{4}E\{|p_{\mathrm{rev}}(\mathbf{r}_0)|^2\cos^2(\phi(\mathbf{r}_0))\}$$
$$= \frac{|Q|^2}{8}E\{|p_{\mathrm{rev}}(\mathbf{r}_0)|^2\}, \quad (8)$$

where ϕ is the phase angle of $p_{\mathrm{rev}}(\mathbf{r}_0)$, which is a random variable uniformly distributed between 0 and 2π. The expectation of the mean square reverberant pressure can be approximated by its spatial average, which is related to the actual sound power emitted by the monopole and the total absorption area of the room by the energy balance equation,[12]

$$E\{|p_{\mathrm{rev}}|^2\} \simeq \langle|p_{\mathrm{rev}}|^2\rangle = \frac{8\rho c}{A}P_a, \quad (9)$$

where $\langle\ \rangle$ indicates a spatial average. Equation (8) now becomes

$$\sigma^2\{P_a\} \simeq \frac{|Q|^2}{8}\frac{8\rho c}{A}P_a \simeq E^2\{P_a\}\frac{8\pi}{Ak^2} = \frac{E^2\{P_a\}}{M_s}, \quad (10)$$

where use has been made of Eq. (7) and P_a has been approximated by its expectation. Normalizing with the free-field sound power finally gives

$$\varepsilon\{P_a\} \simeq \frac{1}{k}\sqrt{\frac{8\pi}{A}} = \frac{1}{\sqrt{M_s}}. \quad (11)$$

It is interesting to note that this expression is identical to Davy's [Eq. (4)] except for a constant factor, and it is quite surprising that the modal overlap enters into a theory that does not even operate with the concept of a mode. The expression can, of course, easily be extended so as to include the effect of 1D and 2D modes in a rectangular room. Note, however, that no assumptions about the shape (or size) of the room have been made.

Expression (11) does not take account of the phenomenon of "weak Anderson localization," also known as coherent backscattering,[7,14,15] according to which there is a concentration of the *reverberant* sound field at the source position. Such a phenomenon was already predicted on the basis of Poisson statistics in Lyon's paper.[1] According to Langley and Cotoni,[7] the now generally accepted assumption of Gaussian orthogonal ensemble statistics leads to a "concentration factor" for the mean square reverberant field that approaches 3 for $M_s \to 0$ and 2 for $M_s \to \infty$. This can be expected to modify Eq. (11) to

$$\varepsilon\{P_a\} = \sqrt{\frac{F(M_s)}{M_s}}, \quad (12)$$

where the concentration factor F is a function that goes smoothly from 3 to 2 as the modal overlap increases. It is worth noting that Eqs. (4) and (12) are *identical* if $K=3$ (as suggested in Refs. 3 and 8) and $F=2$.

TABLE I. Reverberation rooms used in the experiments.

Description of room	Volume (m^3)	Reverberation time at 500 Hz (s)	Schroeder frequency (Hz)
Small lightly damped room	40	2.3	460
Small heavily damped room	40	0.8	290
Large reverberation room	245	5.7	300
Very large reverberation room	500	5.8	220
Large strongly damped hall	650	1.1	90

IV. EXPERIMENTAL AND NUMERICAL RESULTS

To examine the validity of the established theory and the alternative simpler one, some experiments have been carried out in different reverberation rooms, but since it is impossible in practice to determine statistical properties associated with an ensemble of rooms the experiments are supplemented by finite element calculations.

A. Experimental results

A "volume velocity source," Brüel & Kjær (B&K) 4295 (a tube with two matched quarter-inch microphones driven by a loudspeaker), can also be used for measuring the radiation impedance of the small opening of the tube, which is related to the frequency response between the two microphone signals (see Appendix A). This method has been used for measuring the radiation impedance at 24 positions in four different rooms, one of which in two different conditions (see Table I). Figure 1 shows the reverberation time of the rooms measured in one-third octave bands with a B&K "PULSE" analyzer using interrupted noise.

The frequency response between the microphone signals from the volume velocity source was measured with the same analyzer in fast Fourier transform mode with 6400

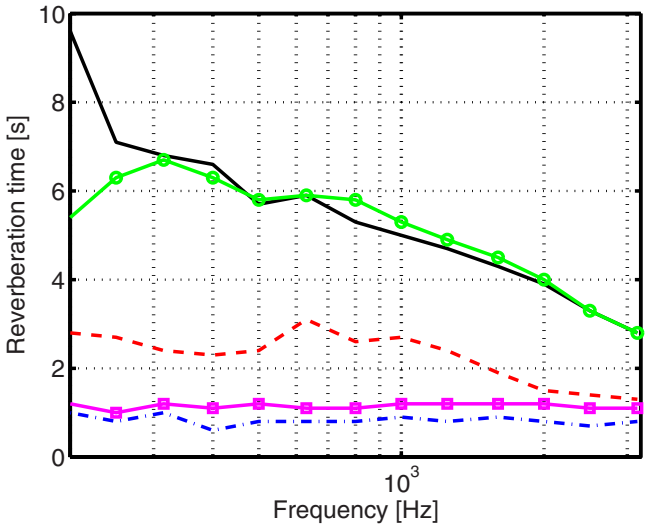


FIG. 1. (Color online) Reverberation time of the test rooms. Solid line: large reverberation room; solid line with circles: very large reverberation room; dashed line: small lightly damped room; solid line with squares: large heavily damped hall; dash-dotted line: small heavily damped room.

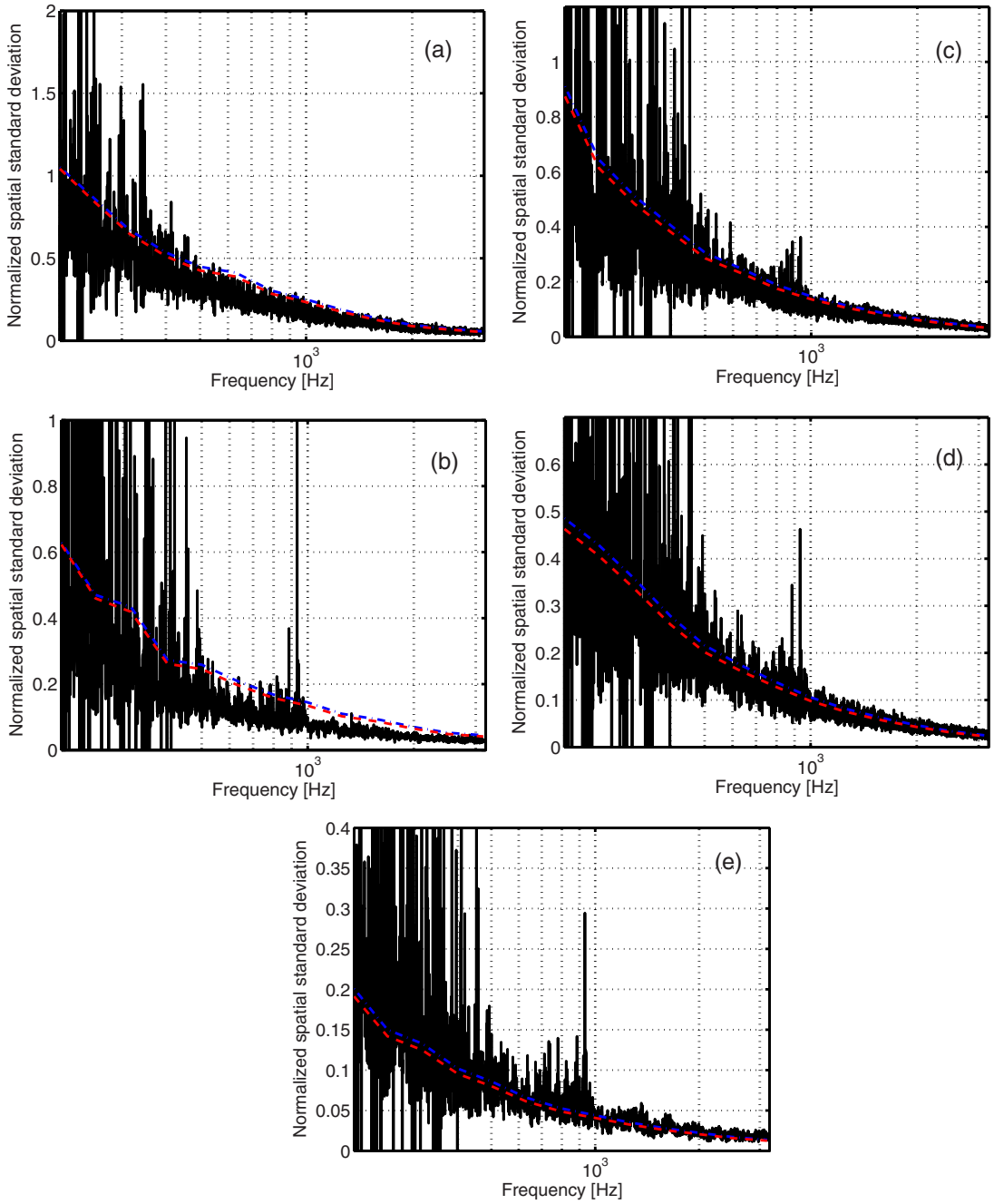


FIG. 2. (Color online) Normalized spatial standard deviation of sound power output in (a) small lightly damped small room, (b) small heavily damped room, (c) large reverberation room, (d) very large reverberation room, and (e) large heavily damped hall. Solid line: measured standard deviation; dash-dotted line: prediction based on Eq. (4); dashed line: prediction based on Eq. (12) with $F=2$.

spectral lines and a resolution of 0.5 Hz. The source was driven with synchronized pseudorandom noise generated by the PULSE analyzer, and a uniform time window was used. This corresponds to measuring at 6400 independent discrete frequencies. A similar technique has been used by Baade and Maling[16] in "qualification" of reverberation rooms. Measuring the radiation impedance of the volume velocity source directly is extremely efficient compared with measuring the sound power output with a conventional method using many source and receiver positions, but it should be mentioned that this technique makes heavy demands on the equipment as reflected in many (meaningless) negative estimates at low frequencies (see Appendix A).

Figure 2 compares the normalized spatial standard deviation of the real part of the radiation impedance observed in the five rooms with predictions based on Davy's Eq. (4) and the new expression, Eq. (12), with $F=2$ because the modal overlap is large except at very low frequencies. Since all the rooms are essentially rectangular, Eqs. (4) and (12) have been modified to take account also of 2D and 1D modes, although the effect of this modification is almost negligible. It is interesting that the two fundamentally different theories give practically identical predictions. At low frequencies, the measured spatial standard deviations fluctuate enormously with the frequency. This is not unexpected since the spatial standard deviation at low modal overlap depends

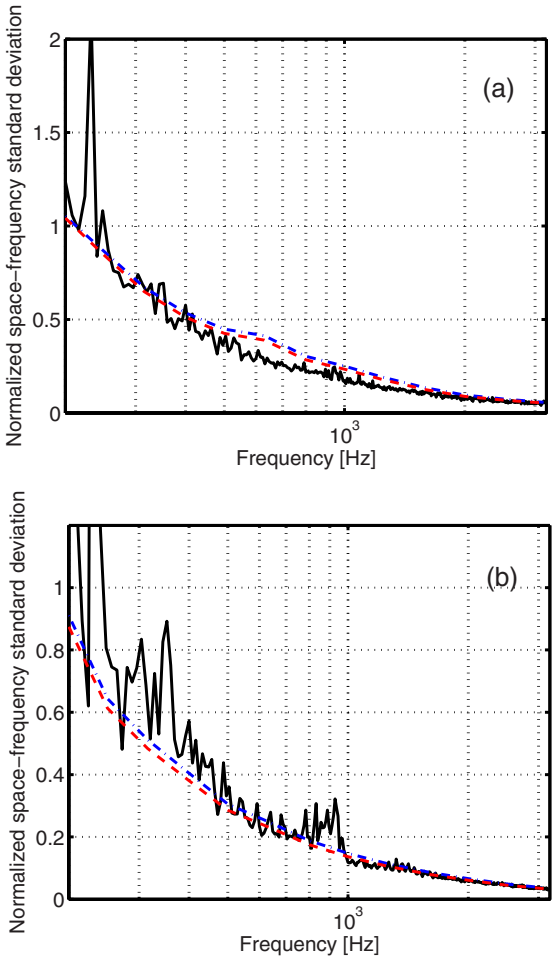


FIG. 3. (Color online) Normalized space-frequency standard deviation of sound power output in (a) small lightly damped room and (b) large reverberation room. Solid line: measured standard deviation; dash-dotted line: prediction based on Eq. (4); dashed line: prediction based on Eq. (12) with $F=2$.

strongly on whether one single mode dominates or whether several modes contribute to the sound field. The spatial standard deviation tends to be slightly lower than the predicted ensemble standard deviation except in the large hall, but it certainly has the same tendency as the predictions in all cases.

Figure 3 shows the corresponding space-frequency standard deviation in two of the five rooms, calculated from variations in the same data with respect not only to source positions but also to 16 neighboring frequencies (corresponding to a band of 8 Hz). The latter is an attempt to estimate the ensemble standard deviation. There is fair, if not perfect, agreement between predictions and experimental results.

B. Finite element calculations

An alternative approach involves calculating the ensemble and spatial sound power variance numerically using the finite element method (FEM). With a FEM model of a room enclosing a point source, the sound power can be calculated simply by performing the integration of the normal component of the sound intensity along the boundaries. This integration can readily be performed by means of functions available in most FEM packages. If the position of the point source is varied, the spatial variance in the room can be obtained. If, in addition, the calculation is carried out in a set of rooms with different shapes, and thus with different modal overlaps, the ensemble variance of sound power can be achieved by computing the variance of sound power obtained for all source positions in all tested rooms. However, the calculation of the ensemble variance must be performed directly in terms of the modal overlap. Since this quantity depends on the total absorption area of the room, or its surface area and wall impedance [see Eq. (2)], the modal overlaps of the various rooms may differ. If one would try to compute the variance of all points for a certain value of the modal overlap, only values of a single room would be available. The solution chosen here is to divide the modal overlap axis into bands. Performing the calculation in such modal overlap intervals makes it possible to estimate the ensemble variance, provided that there are enough results from different rooms in the intervals.

Two sets of calculations were carried out: three-dimensional (3D) calculations that can be compared directly with the experimental data and 2D calculations that make it possible to increase the frequency range. The equations for 2D are given in Appendix B. For simplicity, the rooms were rectangular. All calculations were made using the FEM software package COMSOL 3.4, with which one can specify excitation by a monopole at a given position. To guarantee a low numerical pollution, the quantity $kh/2p$ (in which h is the maximum size of the elements and p is their order) was kept less than 0.5.[17] In both cases, the sound source was placed randomly in the room, provided that it was at least 0.4 m away from the walls. For the 3D calculations the room dimensions were changed from $2\times5\times4.27$ to $4\times3\times3.56$ m^3, and for the 2D calculations the dimensions of the "room" were changed from 2×7.2 to 4×3.6 m^2. In all cases, the walls were locally reacting with an impedance of 87 000 Pa s/m. The 3D calculations were carried out from 50 to 300 Hz with a frequency step of 1 Hz, and the ensemble variance was calculated in modal overlap bands with a width of 0.09. The 2D calculations were carried out in the same frequency range with a frequency step of 12.5 Hz and modal overlap intervals of 0.02.

Figure 4(a) compares the calculated ensemble standard deviation of the sound power with respect to 50 different 3D rooms and 50 source positions with predictions calculated using Eqs. (4) and (12), and Fig. 4(b) does the same for the 2D case, in both cases with $F=2$ as well as with $F=3$ in Eq. (12). In the 3D case, the two theories lead to practically identical predictions with $F=2$, but the agreement with the FEM results is better, in fact almost perfect, with $F=3$. This agrees with Langley and Cotoni's[7] prediction of a concentration factor of 3 at low modal overlap. In the 2D case, Eq. (4) underestimates and Eq. (12) overestimates the observed standard deviations with $F=3$, whereas Eq. (12) agrees extremely well with the FEM results with $F=2$.

V. DISCUSSION

It should be emphasized that the underlying theories of Eqs. (4) and (12) are quite different. Davy's model is based

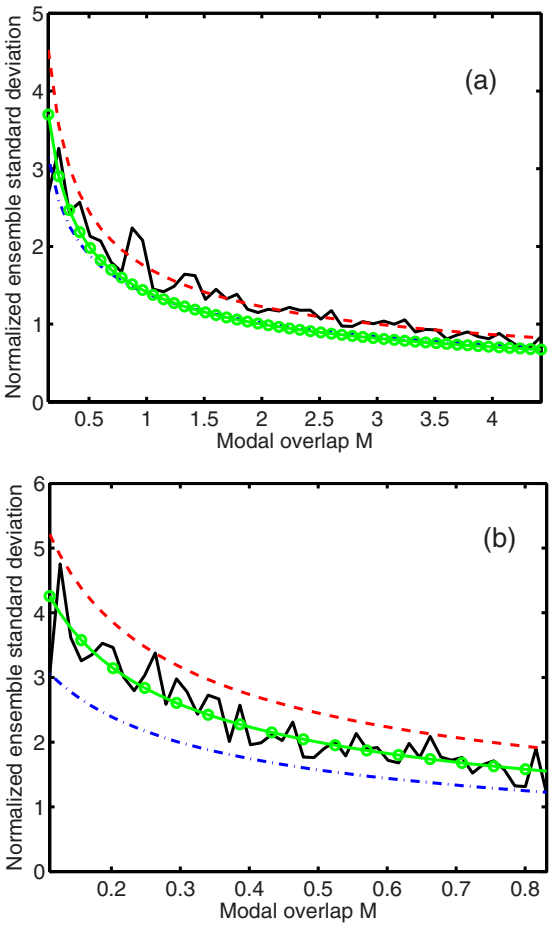


FIG. 4. (Color online) Normalized ensemble standard deviation of sound power output in (a) 3D rooms and (b) 2D rooms. Solid line: FEM results; dash-dotted line: prediction based on Eq. (4); solid line with circles: prediction based on Eq. (12) with $F=2$; dashed line: prediction based on Eq. (12) with $F=3$.

on the analytical Green's function in a rectangular room, makes use of the concept of impedance, and requires knowledge of mode shapes and the modal density and distribution. The alternative theory does not even recognize the concept of modes, and the losses of the rooms are described in terms of the absorption area. Thus the almost perfect agreement between the predictions (in the 3D case) is surprising. Note, however, that the factor of F has been derived in the literature using a modal approach.[7,14] The free wave model cannot possibly take account of the coherent backscattering effect since the waves might be generated by independent sources.

Both the established theory and the alternative one should be taken with a grain of salt at *very* low frequencies. For example, it is obvious that at extremely low frequencies, well below the lowest modal frequency, both theories predict a huge variance, whereas in reality the spatial source position variance is zero and the variation between rooms is only related to their different volumes. The spatial source position variance is also limited in the single mode case (19/8, 5/4, and 1/2 for 3D, 2D, and 1D modes in a rectangular room, respectively[18]).

At somewhat higher frequencies but below the Schroeder frequency, the spatial variance is strongly affected by whether the frequency coincides with a modal frequency. However, above the Schroeder frequency where the modal overlap is high, one would expect the same statistics with respect to room, position, and frequency. In other words, merely varying the source position can be expected to lead to the full ensemble variance in this frequency range because many modes are excited at all frequencies, and this ergodicity seems to be confirmed by the experimental results, although Schroeder's original "large room frequency," which corresponds to twice the established Schroeder frequency,[19] is perhaps more adequate.

VI. CONCLUSION

Experimental and numerical results confirm the well-known observation that there is a substantial uncertainty in sound power measurements of pure-tone sources in non-anechoic rooms at low frequencies. The associated standard deviation is inversely proportional to the square root of the modal overlap, that is, essentially proportional to the square root of the ratio of the reverberation time to the volume of the room and inversely proportional to the frequency.

Experimental results obtained in four very different rooms, one of which in two different damping conditions, confirm that the sound power output of the monopole exhibits the same statistical fluctuations with respect to changes in the source position and changes in the frequency somewhat above the Schroeder frequency.

An extremely simple, very general model based on sound waves arriving with random phases from random directions has, surprisingly, been found to give predictions of the ensemble standard deviation of the sound power of a monopole in almost perfect agreement with the predictions of the established, far more complicated theory. These predictions are in fairly good agreement with experimental observations and in excellent agreement with the results of finite element calculations in three dimensions. The alternative model indirectly confirms the phenomenon known as weak Anderson localization in reverberation rooms and seems also to confirm that the concentration factor, which describes the local increase in the reverberant sound energy at the source position, is 3 at low modal overlap and 2 at higher modal overlap.

In 2D rooms, the results are slightly different. The general tendencies are the same, but in this case there is a systematic difference between the two theories, and the alternative one is in best agreement with the finite element calculations if the assumed concentration factor is 2.

ACKNOWLEDGMENTS

The first author would like to thank John Davy and Richard Weaver for stimulating discussions during the Acoustics'08 meeting in Paris in June/July 2008.

APPENDIX A: EXPERIMENTAL TECHNIQUE

A method of measuring the volume velocity of an "experimental monopole" has been described in Refs. 20 and 21. The method is based on a tube with two matched quarter-inch microphones driven by a loudspeaker at the other end,

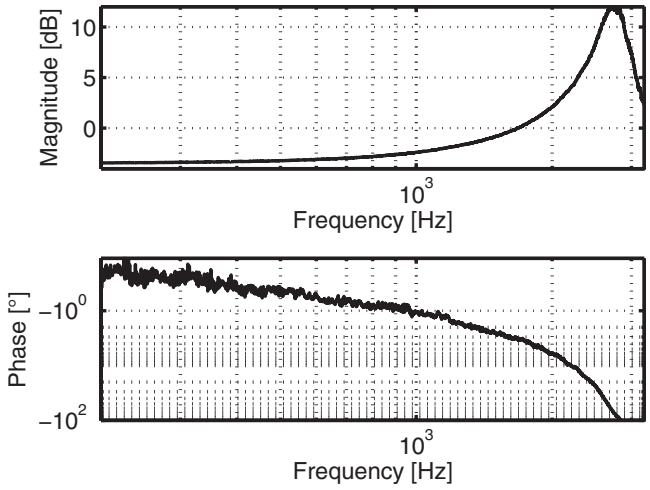


FIG. 5. Magnitude and phase of a frequency response in the tube, H_{AB}.

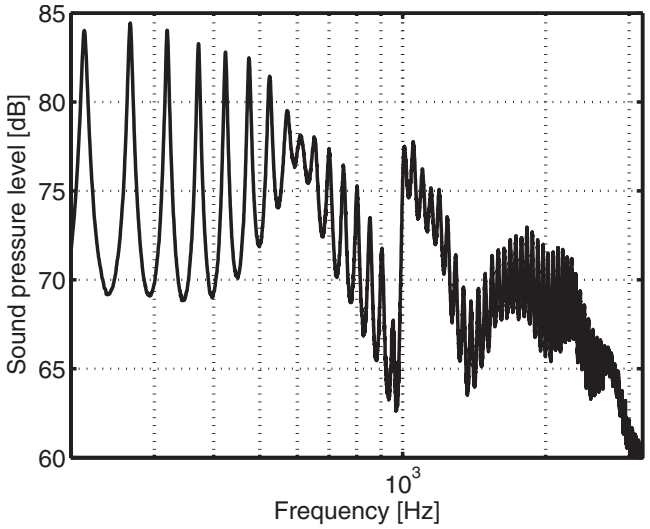


FIG. 6. Power spectrum of the sound pressure measured near the opening of the tube.

and this device, B&K 4295, can also be used for measuring the radiation impedance of the small opening of the tube. In the frequency range where it can be assumed that only plane waves propagate in the tube, the volume velocity at the opening of the tube can readily be shown to be

$$Q = \frac{S_t}{\rho c} \cdot \frac{p_A \cos kl - p_B \cos[k(l+\Delta l)]}{j \sin k\Delta l}, \tag{A1}$$

where S_t is the cross-sectional area of the tube, p_A and p_B are the two pressure signals, Δl is the distance between the microphones, and l is the distance between the opening and the microphone nearest the opening. The sound pressure at the opening is

$$p = \frac{-p_A \sin kl + p_B \sin[k(l+\Delta l)]}{\sin k\Delta l}. \tag{A2}$$

It now follows that the real part of the radiation impedance is

$$\mathrm{Re}\{Z_r\} = \mathrm{Re}\left\{\frac{p}{Q}\right\} = -\frac{\rho c}{S_t} \frac{\mathrm{Im}\{p_A^* p_B\}\sin(k\Delta l)}{|p_A|^2 \cos^2(kl) + |p_B|^2 \cos^2[k(l+\Delta l)] - 2\,\mathrm{Re}\{p_A^* p_B\}\cos(kl)\cos[k(l+\Delta l)]}. \tag{A3}$$

Expressed in terms of the frequency response between the two pressure signals, this becomes

$$\mathrm{Re}\{Z_r\} = -\frac{\rho c}{S_t} \frac{\mathrm{Im}\{H_{AB}\}\sin(k\Delta l)}{\cos^2(kl) + |H_{AB}|^2 \cos^2[k(l+\Delta l)] - 2\,\mathrm{Re}\{H_{AB}\}\cos(kl)\cos[k(l+\Delta l)]}. \tag{A4}$$

The measurement is not without problems, and at low frequencies some estimates of the real part of the radiation impedance are negative, which is why it was chosen to present the results only from 200 Hz and upward. The pressure-intensity index, a quantity well known from the literature on sound intensity measurements with conventional sound intensity probes based on two matched pressure microphones,[22] can easily be calculated from the frequency response between the microphone signals,[23]

$$10 \log\left(\frac{p_{\mathrm{mean}}^2/(\rho c)}{I}\right) = 10 \log\left(\frac{(1 + |H_{AB}|^2 + 2\,\mathrm{Re}\{H_{AB}\})/4}{-\mathrm{Im}\{H_{AB}\}/(k\Delta l)}\right), \tag{A5}$$

and is about 13 dB below 1 kHz, indicating rather harsh sound field conditions. However, the microphones of B&K 4295 are matched so well that this is not a very serious problem, and an attempt to compensate for possible phase and amplitude mismatch (using the "sensor-switching technique" suggested by Chung and Blaser[24]) did not improve conditions.

Figure 5 shows an example of a frequency response (H_{AB}) measured in the large hall. Note that the phase angle of this function varies between about $-0.2°$ at 200 Hz and about $-1°$ at 1 kHz. Inspection of Eq. (A4) shows that the information about the real part of the radiation impedance (or the emitted sound power) is associated with the imaginary part of the frequency response, which is obviously very small below 1 kHz, and the information about how the emitted sound power varies from position to position is hidden in minute fluctuations in this very small imaginary part of the frequency response. There are strong resonances in the loudspeaker cavity and tube, as can be seen in Fig. 6, which

shows the spectrum of the microphone signal nearest the opening of the tube. These resonances should obviously not affect the radiation impedance at the outlet of the tube. However, it is believed that the most serious source of error in the experimental results presented in Sec. IV is due to numerical problems in the signal processing caused by the imaginary part of the frequency response being several hundred times smaller than the real part. Thus the resonances may nevertheless have had some effects on the *estimated* radiation impedance.

APPENDIX B: SOUND POWER IN 2D ROOMS

By analogy with Eq. (5), a 2D diffuse sound field can be modeled as an infinite sum of plane waves coming from all angles of incidence (from 0 to 2π) and having random phases. Such a sound field might conceivably be generated by an infinite pulsating line source in an infinitely long room. The sound intensity that is incident on the walls of such a room is

$$I'_{\mathrm{inc}} = \frac{1}{2\pi}\int_{-\pi/2}^{\pi/2} \frac{E\{|p_{\mathrm{rev}}|^2\}}{2\rho c}\cos\,\theta d\theta = \frac{E\{|p_{\mathrm{rev}}|^2\}}{2\pi\rho c}, \tag{B1}$$

and thus the sound power per unit length emitted by the source and absorbed by the walls is

$$P'_a \simeq \frac{E\{|p_{\mathrm{rev}}|^2\}}{2\pi\rho c}A', \tag{B2}$$

where A' is the absorption area of the walls of the room per unit length. On average, the sound power emitted by the line source equals its free-field sound power,[25]

$$P'_a = \frac{|Q'|^2}{2}\mathrm{Re}\{Z'_r\} = \frac{\rho c k|Q'|^2}{8}, \tag{B3}$$

where Q' is the volume velocity per unit length and Z'_r is the radiation impedance. However, because of the reverberant, diffuse sound field, the sound power output varies from position to position and from room to room. The corresponding variance is

$$\begin{aligned}\sigma^2\{P'_a\} &= \frac{|Q'|^4}{4}\sigma^2\{\mathrm{Re}\{Z'_r\}\} = \frac{|Q'|^4}{4}\sigma^2\left\{\mathrm{Re}\left\{\frac{p_{\mathrm{rev}}}{Q'e^{j\omega t}}\right\}\right\}\\ &= \frac{|Q'|^2}{4}E\{|p_{\mathrm{rev}}|^2\cos^2\,\varphi\} = \frac{|Q'|^2}{8}E\{|p_{\mathrm{rev}}|^2\}\\ &\simeq \frac{|Q'|^2}{8}\frac{2\pi\rho c}{A'}E\{P'_a\} = E^2\{P'_a\}\frac{2\pi}{A'k}.\end{aligned} \tag{B4}$$

If the source is suddenly turned off, the sound energy in the room decays with a time constant that follows from simple energy balance considerations,

$$\tau' = \frac{E'_a}{P'_a} = \frac{S'}{2L'c\,\mathrm{Re}\{\beta'\}} = \frac{\pi S'}{cA'}, \tag{B5}$$

where $S' = l_x \cdot l_y$ is the cross-sectional area of the infinitely long room, $L' = 2(l_x + l_y)$, and β' is the normalized wall admittance. The modal density in modes per hertz is

$$n'(f) = \frac{2\pi S' f}{c^2}, \tag{B6}$$

and the statistical modal overlap is

$$M'_s = \frac{n'(f)}{2\tau'} = \frac{A'k}{2\pi}. \tag{B7}$$

Combining Eqs. (B4) and (B7) finally shows that the normalized standard deviation of the emitted sound power has exactly the same form in 2D as in 3D,

$$\varepsilon\{P'_a\} = \frac{1}{\sqrt{M'_s}}. \tag{B8}$$

This expression should be modified in the same way as Eq. (11) because of the effect of weak Anderson localization.

In 2D, Davy's Eq. (4) becomes

$$\varepsilon\{P'_a\} = \frac{1}{\sqrt{M_s}}\left(\left(\frac{3}{2}\right)^2 - 1\right)^{1/2}. \tag{B9}$$

One can easily extend Eqs. (B8) and (B9) so as to take account of 1D modes.

[1] R. H. Lyon, "Statistical analysis of power injection and response averaging in structures and rooms," J. Acoust. Soc. Am. **45**, 545–565 (1969).
[2] J. L. Davy, "The relative variance of the transmission function of a reverberation room," J. Sound Vib. **77**, 455–479 (1981).
[3] R. L. Weaver, "On the ensemble variance of reverberation room transfer functions, the effect of spectral rigidity," J. Sound Vib. **130**, 487–491 (1989).
[4] R. V. Waterhouse, "Estimation of monopole power radiated in a reverberation chamber," J. Acoust. Soc. Am. **64**, 1443–1446 (1978).
[5] J. L. Davy, "Improvements to formulae for the ensemble relative variance of random noise in a reverberation room," J. Sound Vib. **115**, 145–161 (1987).
[6] T. A. Brody, J. Flores, J. B. French, P. A. Mello, A. Pandey, and S. S. M. Wong, "Random matrix physics: Spectrum and strength fluctuations," Rev. Mod. Phys. **53**, 385–479 (1981).
[7] R. S. Langley and V. Cotoni, "The ensemble statistics of the vibrational energy density of a random system subjected to single point harmonic excitation," J. Acoust. Soc. Am. **118**, 3064–3076 (2005).
[8] O. I. Lobkis, R. L. Weaver, and I. Rozhkov, "Power variances and decay curvature in a reverberant system," J. Sound Vib. **237**, 281–302 (2000).
[9] J. L. Davy, "The distribution of modal frequencies in a reverberation room," in Proceedings of Inter-Noise 90, Gothenburg, Sweden (1990).
[10] J. L. Davy, "The variance of pure tone reverberant sound power measurements," in Proceedings of the Fifth International Congress on Sound and Vibration, Adelaide, Australia, (1997).
[11] F. Jacobsen, "The diffuse sound field," Ph.D. thesis, The Acoustics Laboratory, Technical University of Denmark, Denmark (1979).
[12] D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981), Secs. 6.3 and 6.1.
[13] R. V. Waterhouse, "Statistical properties of reverberant sound fields," J. Acoust. Soc. Am. **43**, 1436–1444 (1968).
[14] R. L. Weaver and J. Burkhardt, "Weak Anderson localization and enhanced backscatter in reverberation rooms and quantum dots," J. Acoust. Soc. Am. **96**, 3186–3190 (1994).
[15] G. Tanner and N. Søndergaard, "Wave chaos in acoustics and elasticity," J. Phys. A: Math. Theor. **40**, R443–R509 (2007).
[16] P. K. Baade and G. C. Maling, "Reverberation room qualification using multitone signals," Noise Control Eng. J. **46**, 23–28 (1998).
[17] F. Ihlenburg, *Applied Mathematical Sciences* (Springer-Verlag, New York, 1998), Vol. **132**.
[18] F. Jacobsen, "Sound power determination in reverberation rooms: A normal mode analysis," Ph.D. thesis, The Acoustics Laboratory, Technical University of Denmark, Denmark (1979).

[19]M. Schröder, "Die statistichen Parameter der Frequenzkurven von grossen Räumen (Statistical parameters of the frequency response curves of large rooms)," Acustica **4**, 594–600 (1954).
[20]S. Gade, N. Møller, J. Hald, and L. Alkestrup, "The use of volume velocity source in transfer measurements," in Proceedings of Inter-Noise 2004, Prague, Czech Republic (2004).
[21]Y. Luan and F. Jacobsen, "A method of measuring the Green's function in an enclosure," J. Acoust. Soc. Am. **123**, 4044–4046 (2008).
[22]F. J. Fahy, *Sound Intensity*, 2nd ed. (E & FN Spon, London, 1995).
[23]F. Jacobsen, "Sound field indicators: Useful tools," Noise Control Eng. J. **35**, 37–46 (1990).
[24]J. Y. Chung and D. A. Blaser, "Transfer function method of measuring in-duct acoustic properties. I. Theory," J. Acoust. Soc. Am. **68**, 907–913 (1980).
[25]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), Sec. 7.3.

Acoustic response of a rigid-frame porous medium plate with a periodic set of inclusions

J.-P. Groby[a)]
CMAP, CNRS/École Polytechnique, UMR7641, F-91128 Palaiseau Cedex, France

A. Wirgin[b)]
Laboratoire de Mécanique et d'Acoustique, CNRS, UPR7051, Marseille, France

L. De Ryck[c)] and W. Lauriks
Laboratory of Acoustics and Thermal Physics, KULeuven, Heverlee, Belgium

R. P. Gilbert
Department of Mathematical Sciences, University of Delaware, Newark, Delaware 19716

Y. S. Xu
Department of Mathematics, University of Louisville, Louisville, Kentucky 40292

(Received 2 March 2009; revised 2 June 2009; accepted 2 June 2009)

The acoustic response of a rigid-frame porous plate with a periodic set of inclusions is investigated by a multipole method. The acoustic properties, in particular, the absorption, of such a structure are then derived and studied. Numerical results together with a modal analysis show that the addition of a periodic set of high-contrast inclusions leads to the excitation of the modes of the plate and to a large increase in the acoustic absorption. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3158936]

PACS number(s): 43.55.Ev, 43.20.Fn, 43.20.Ks, 43.20.Gp [NX] Pages: 685–693

I. INTRODUCTION

This work was initially motivated by the design problem connected with the determination of the optimal profile of a continuous and/or discontinuous spatial distribution of the material/geometric properties of porous materials for the absorption of sound. Porous materials (foam) suffer from a lack of absorption particularly at low frequency, when compared to its value at higher frequency, and for normal incidence. The usual way to solve this problem is by multi-layering.[1–3] The purpose of the present article is to investigate an alternative to multi-layering by embedding a periodic set of inclusions, whose size is not small compared with the wavelength, in an otherwise macroscopically-homogeneous porous plate whose thickness and weight are relatively small (i.e., the principal constraints in the design of acoustic absorbing materials).

Acoustic wave propagation in porous materials was mainly studied in order to deal with sound absorption, material property characterization, etc. Homogeneous porous materials are well described by the first work of Biot[4,5] and later contributions.[6–8] On the other hand, the equation that describes acoustic wave propagation in a macroscopically-inhomogeneous rigid-frame porous medium was only recently derived in Ref. 9 from the alternative formulation of Biot's theory.[5] The latter equation could eventually offer an alternative to multi-layering in the sense that it can be applied to the design of (e.g., functionally gradient) rigid-frame media with continuously-varying properties.

The influence of the addition of a volumic heterogeneity on absorption and transmission of porous plates was previously studied by means of the homogenization procedure. In Ref. 10, the authors considered the reflection of a plane acoustic wave by a porous plate that presents a periodic set of pits oriented along the direction of propagation. The medium is homogenized, generalized expressions for the complex bulk modulus and dynamic permeability of double-porosity media are presented, and its behavior is described in Ref. 11. This leads to a drastic increase in the absorption coefficient at low frequency. In Ref. 12, the authors considered the transmission of an acoustic wave through a porous medium in which a set of randomly-arranged metallic rods is embedded. This medium is converted, by a procedure called *ISA*β, into an equivalent homogeneous medium, which exhibits decreased transmission and increased absorption. Herein, the diffraction of a plane wave by a porous plate in which is embedded a periodic set of porous or high-contrast (Neumann type condition) inclusions is studied with the help of the so-called multipole-method.[13–15] The field scattered by the inclusions is fully accounted-for so that a complete description of the effects, in terms of increase in the absorption, is made possible.

Periodic arrangements of either surface irregularities or volume heterogeneities usually lead to energy entrapment either at the surface or inside the structure, this being strongly linked to mode excitation and to an increase in the absorption coefficient (first noticed by Wood[16] and partially ex-

[a)]Electronic mail: jeanphilippe.groby@fys.kuleuven.be
[b)]Present address: Laboratory of Acoustics and Thermal Physics, Celestijnenlaan 200D, B-3001 Leuven, Belgium.
[c)]Present address: LMS International NV, 3001 Leuven, Belgium.

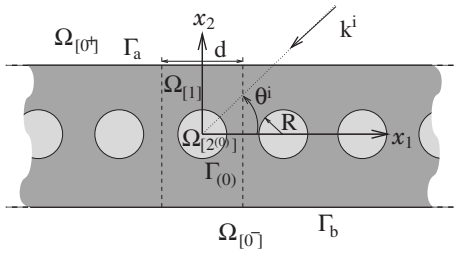


FIG. 1. Sagittal plane representation of the configuration of plane wave solicitation of a d-periodic porous fluid-like plate with fluid-like inclusions (of radius R).

plained by Cutler[17]). The particular properties of such structures have been studied in mechanics, with application to composite materials,[18–20] in optics initially motivated by the collection of solar energy[21,22] with applications to photonic crystals,[23,24] in electromagnetics with application to so-called left-handed materials,[25] and in geophysics for the study of the "city-site" effect.[26,27] The properties of such structures are now studied to create band-gaps for elastic or acoustic waves (phononic/sonic crystals[28–30]), and have been used for the design of sound absorbing or porous materials.[31–34]

II. FORMULATION OF THE PROBLEM

A. Description of the configuration

Both the incident plane acoustic wave and the plate are assumed to be invariant with respect to the Cartesian coordinate x_3. A sagittal x_1-x_2 plane view of the two-dimensional scattering problem is given in Fig. 1.

Before the introduction of the cylindrical inclusions, the plate is made of a porous material (e.g., a foam) which is modeled (by homogenization) as a (macroscopically-homogeneous) equivalent fluid $M^{[1]}$. Another equivalent fluid medium $M^{[2]}$ occupies each cylindrical inclusion. In the sagittal plane, the jth cylinder is the circular disk $\Omega_{[2^{(j)}]}$. Two subspaces $\Omega_{[1^\pm]}$ are also created, respectively, corresponding to the upper and lower parts of the plate containing the inclusions. The host medium $M^{[0]}$ occupying the two half spaces $\Omega_{[0^\pm]}$ is air. Thus, the plate is macroscopically-inhomogeneous, the heterogeneity being periodic in the x_1 direction with period d.

The upper and lower flat, mutually-parallel boundaries of the plate are Γ_a and Γ_b in the sagittal plane. The x_2 coordinates of these lines are a and b, the thickness L of the plate being $L=a-b$. The circular boundary of $\Omega_{[2^{(j)}]}$ is $\Gamma_{(j)}$. The center of the $j=0$ disk is at the origin O of the laboratory system $Ox_1x_2x_3$. The union of $\Omega_{[0^+]}$ and $\Omega_{[0^-]}$ is denoted by $\Omega_{[0]}$.

The wavevector $\mathbf{k}^i$ of the incident plane wave lies in the sagittal plane and the angle of incidence is θ^i measured counterclockwise from the positive x_1 axis.

B. Wave equations

Total pressure, wavenumber, and wave speed are designated by the generic symbols p, k, and c respectively, with $p=p^{[j]}$, $k=k^{[j]}=\omega/c^{[j]}$ in $\Omega_{[j]}$, $j=0^\pm,1,2^{(j)}$.

Rather than solve directly for the pressure $p(\mathbf{x},t)$ [with $\mathbf{x}=(x_1,x_2)$], $\tilde{p}(\mathbf{x},\omega)$ is preferred, related to $p(\mathbf{x},t)$ by the Fourier transform $p(\mathbf{x},t)=\int_{-\infty}^{\infty}\tilde{p}(\mathbf{x},\omega)e^{-i\omega t}d\omega$. Henceforth, the ω is dropped in $\tilde{p}(\mathbf{x},\omega)$ so as to designate the latter by $\tilde{p}(\mathbf{x})$.

The effective compressibility and density of medium $M^{[1]}$, linked to the sound speed through $c^{[1]}=\sqrt{1/(K^{[1]}\rho^{[1]})}$, are[8,9]

$$\frac{1}{K^{[1]}}=\frac{\gamma P_0}{\phi\left(\gamma-(\gamma-1)\left(1+i\frac{\omega_c}{\mathrm{Pr}\omega}G(\mathrm{Pr}\omega)\right)^{-1}\right)},$$

$$\rho^{[1]}=\frac{\rho_f\alpha_\infty}{\phi}\left(1+i\frac{\omega_c}{\omega}F(\omega)\right), \tag{1}$$

wherein $\omega_c=\sigma\phi/\rho_f\alpha_\infty$ is the Biot characteristic frequency, γ is the specific heat ratio, P_0 is the atmospheric pressure, Pr is the Prandtl number, ρ_f is the density of the fluid in the (interconnected) pores, ϕ is the porosity, α_∞ is the tortuosity, and σ is the flow resistivity. The correction functions $G(\mathrm{Pr}\omega)$ (Ref. 35) and $F(\omega)$ (Ref. 6) are given by

$$G(\mathrm{Pr}\omega)=\sqrt{1-4i\frac{\eta\rho_f\alpha_\infty^2}{\sigma^2\phi^2\Lambda'^2}\mathrm{Pr}\omega},$$

$$F(\omega)=\sqrt{1-4i\frac{\eta\rho_f\alpha_\infty^2}{\sigma^2\phi^2\Lambda^2}\omega}, \tag{2}$$

wherein η is the viscosity of the fluid, Λ' is the thermal characteristic length, and Λ is the viscous characteristic length.

The incident wave propagates in $\Omega_{[0^+]}$ and is expressed by $\tilde{p}^i(\mathbf{x})=A^ie^{i(k_1^ix_1-k_2^i(x_2-a))}$, wherein $k_1^i=-k^{[0]}\cos\theta^i$, $k_2^i=k^{[0]}\sin\theta^i$, and $A^i=A^i(\omega)$ is the signal spectrum.

The particular feature of the problem is the transverse periodicity of $\cup_{j\in\mathbb{Z}}\Omega_{[2^{(j)}]}$. The plane wave nature of the incident wave and the periodic nature of $\cup_{j\in\mathbb{Z}}\Omega_{[2^{(j)}]}$ imply the Floquet relation

$$\tilde{p}(x_1+nd,x_2)=\tilde{p}(x_1,x_2)e^{ik_1^ind},\quad \forall\, n\in\mathbb{Z}. \tag{3}$$

Consequently, it suffices to determine the field in the central cell of the plate which includes the disk $\Omega_{[2^{(0)}]}$ in order to obtain the fields, via the Floquet relation, in the other cells. Henceforth, the simplified notation is adopted: $\Omega_{[2]}:=\Omega_{[2^{(0)}]}$, $\Gamma:=\Gamma_{[(0)]}$, and $\tilde{p}^{[2]}=\tilde{p}^{[2^{(0)}]}$.

C. Boundary and radiation conditions

Since $M^{[0]}$ and $M^{[1]}$ are fluid-like, the pressure and the normal velocity are continuous across the interfaces Γ_a and Γ_b:

$$\tilde{p}^{[0^+]}(\mathbf{x})-\tilde{p}^{[1]}(\mathbf{x})=0,\quad \forall\,\mathbf{x}\in\Gamma_a,$$

$$\partial_n\tilde{p}^{[0^+]}(\mathbf{x})/\rho^{[0]}-\partial_n\tilde{p}^{[1]}(\mathbf{x})/\rho^{[1]}=0,\quad \forall\,\mathbf{x}\in\Gamma_a,$$

$$\tilde{p}^{[0^-]}(\mathbf{x})-\tilde{p}^{[1]}(\mathbf{x})=0,\quad \forall\,\mathbf{x}\in\Gamma_b,$$

$$\partial_n \tilde{p}^{[0^-]}(\mathbf{x})/\rho^{[0]} - \partial_n \tilde{p}^{[1]}(\mathbf{x})/\rho^{[0]} = 0, \quad \forall\, \mathbf{x} \in \Gamma_b, \tag{4}$$

wherein $\mathbf{n}$ denotes the generic unit vector normal to a boundary and ∂_n designates the operator $\partial_n = \mathbf{n}\cdot\nabla$.

Since $M^{[1]}$ and $M^{[2]}$ are fluid-like, the pressure and normal velocity are continuous across the interface Γ:

$$\tilde{p}^{[2]}(\mathbf{x}) - \tilde{p}^{[1]}(\mathbf{x}) = 0, \quad \forall\, \mathbf{x} \in \Gamma,$$

$$\partial_n \tilde{p}^{[2]}(\mathbf{x})/\rho^{[2]} - \partial_n \tilde{p}^{[1]}(\mathbf{x})/\rho^{[1]} = 0, \quad \forall\, \mathbf{x} \in \Gamma. \tag{5}$$

When the contrast between the media $M^{[1]}$ and $M^{[2]}$ is high, i.e., high-contrast inclusions, the latter can be approximated as infinitely rigid. In this case, the boundary conditions across the interface Γ no longer depend on the fields inside the inclusions, i.e., on the internal geometry of $\Omega_{[2]}$ and on the material characteristics of $M^{[2]}$ (except its rigid behavior), and reduce to Neumann type boundary conditions:

$$\partial_n \tilde{p}^{[1]}(\mathbf{x}) = 0, \quad \forall\, \mathbf{x} \in \Gamma. \tag{6}$$

The uniqueness of the solution to the forward-scattering problem is assured by the radiation conditions:

$$\tilde{p}^{[0^+]}(\mathbf{x}) - \tilde{p}^{i}(\mathbf{x}) \sim \text{outgoing waves}, \quad \forall\, \mathbf{x} \to \infty,$$

$$\tilde{p}^{[0^-]}(\mathbf{x}) \sim \text{outgoing waves}, \quad \forall\, \mathbf{x} \to \infty. \tag{7}$$

III. FIELD REPRESENTATIONS IN $\Omega_{[0^\pm]}$, $\Omega_{[2]}$, AND $\Omega_{[1^\pm]}$

Separation of variables, the radiation conditions, and the Floquet relation lead to the following representations:

$$\tilde{p}^{[0^+]}(\mathbf{x}) = \sum_{p\in\mathbb{Z}} [A^i e^{-ik_{2p}^{[0]}(x_2-a)}\delta_{p0} + R_p e^{ik_{2p}^{[0]}(x_2-a)}]e^{ik_{1p}x_1}, \tag{8}$$

$$\tilde{p}^{[0^-]}(\mathbf{x}) = \sum_{p\in\mathbb{Z}} T_p e^{i(k_{1p}x_1 - k_{2p}^{[0]}(x_2-b))}, \tag{9}$$

wherein R_p and T_p are, respectively, the reflection and transmission coefficients of the plane wave indiced by p, δ_{p0} the Kronecker symbol, $k_{1p} = k_1^i + 2p\pi/d$, and $k_{2p}^{[0]} = \sqrt{(k^{[0]})^2 - (k_{1p})^2}$, with $\mathrm{Re}(k_{2p}^{[0]}) \geq 0$ and $\mathrm{Im}(k_{2p}^{[0]}) \geq 0$ for $\omega \geq 0$.

The field in the central inclusion, with $[\mathbf{r} = (r,\theta)]$, takes the form

$$\tilde{p}^{[2]}(\mathbf{r}) = \sum_{l\in\mathbb{Z}} C_l J_l(k^{[2]} r) e^{il\theta}, \tag{10}$$

wherein J_l is the lth order Bessel function and C_l are the coefficients of the field scattered inside the cylinder of the unit cell.

It is convenient to use Cartesian coordinates (x_1, x_2) to write the field representation in $\Omega_{[1^\pm]}$. The latter takes the form of the sum (by use of the superposition principle) of the field diffracted by the inclusions and the diffracted field in the plate.[34] Because of the periodic nature of the configuration, the diffracted field in the plate can be written in Cartesian coordinates as

$$\tilde{p}^{[1^\pm]}(\mathbf{x}) = \sum_{p\in\mathbb{Z}} (f_p^{[1]-} e^{-ik_{2p}^{[1]}x_2} + f_p^{[1]+} e^{ik_{2p}^{[1]}x_2}) e^{ik_{1p}x_1} + \sum_{p\in\mathbb{Z}} \sum_{l\in\mathbb{Z}} B_l K_{pl}^{\pm} e^{i(k_{1p}x_1 \pm k_{2p}^{[1]}x_2)}, \tag{11}$$

where $f_p^{[1^\pm]}$ accounts for the diffracted field by the plate, B_l are the coefficients of the field scattered by the cylinder of the unit cell, the signs + and − correspond to $x_2 > R$ and $x_2 < (-R)$, respectively, and $K_{pm}^{\pm} = 2(-i)^m e^{\pm im\theta_p}/dk_{2p}^1$, with θ_p such that $k^{[1]}e^{i\theta_p} = k_{1p} + ik_{2p}^{[1]}$.[13,15]

IV. THE REFLECTED AND TRANSMITTED FIELDS

The method of resolution of the problem is described in Appendix A. Once Eq. (A7) is solved for B_l, $\forall l \in \mathbb{Z}$, expressions for R_p and T_p depending on B_l, $\forall l \in \mathbb{Z}$, are derived from the linear system (A1):

$$R_P = \frac{2A^i i \sin(k_2^{[1]i}L)((\alpha^{[0]i})^2 - (\alpha^{[1]i})^2)}{D_0^i} + \sum_{p\in\mathbb{Z}} \sum_{l\in\mathbb{Z}} \frac{8(-i^l)B_l}{dk_{2p}^{[1]}} \times \frac{\alpha_p^{[1]}(-i\alpha_p^{[0]} \sin(l\theta_p + k_{2p}^{[1]}b) - \alpha_p^{[1]} \cos(l\theta_p + k_{2p}^{[1]}b))}{D_p},$$

$$T_P = \frac{-A^i 4\alpha^{[1]i}\alpha^{[0]i}}{D_0^i} + \sum_{p\in\mathbb{Z}} \sum_{l\in\mathbb{Z}} \frac{8(-i^l)B_l}{dk_{2p}^{[1]}} \times \frac{\alpha_p^{[1]}(i\alpha_p^{[0]} \sin(l\theta_p + k_{2p}^{[1]}a) - \alpha_p^{[1]} \cos(l\theta_p + k_{2p}^{[1]}a))}{D_p}, \tag{12}$$

with

$$D_p = 2i \sin(k_{2p}^{[1]}L)((\alpha_p^{[0]})^2 + (\alpha_p^{[1]})^2) - 4\alpha_p^{[0]}\alpha_p^{[1]} \cos(k_{2p}^{[1]}L), \tag{13}$$

and $D_0^i = D_0$. Introducing the latter equation into Eqs. (8) and (9) leads to the expression of the pressure field in $\Omega_{[0^\pm]}$. In particular, the latter fields (and also in $\Omega_{[1]}$), are the sum of (i) the field in absence of the inclusions (whose expressions are the same as those in Ref. 36) and (ii) the field due to the presence of the inclusions. The latter field, when compared with the Green's function as calculated in Ref. 36 in the case of a line source located in the plate without inclusions, takes the form of the field radiated by induced periodic sources located on Γ. The latter do not add energy to the system, but rather entail a redistribution of the energy in the frequency range of the solicitation and allow the presence of evanescent waves in the ambient media whereby it is possible to excite modes.

Both the method of resolution (including the numerical recipes, Appendix B) and field expression were validated with the help of the authors' finite-element code[37] in the case of a dissipative plate in which a number of dissipative inclusions are embedded, large enough for the situation in the central part of the configuration to be very close to that described by the multipole method in the case of an infinite number of inclusions.

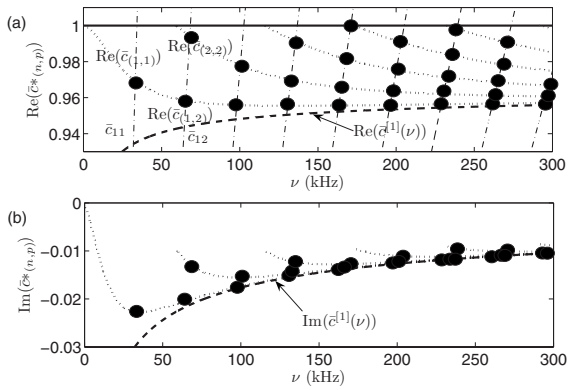


FIG. 2. (a) Real part and (b) imaginary part of the roots of the dispersion relation of a $L=10\times10^{-3}$ m thick porous plate: $(\cdots)\bar{c}^\star(\nu)$, (—) $\bar{c}^{[0]}$, and (- - -) $\bar{c}^{[1]}(\nu)$. Modified modes of the plate and the associated attenuation for period $d=10\times10^{-3}$ m are pointed out by a dot.

V. MODAL ANALYSIS

A. Modal analysis in the absence of inclusions

In the absence of inclusions, the coefficients of the field scattered by the inclusions vanish ($B_l=0,\ \forall l\in\mathbb{Z}$) and all the unknowns that are indexed by p reduce to their value for $p=0$. The problem reduces to the determination of $R=R_p\delta_{p0}$, $T=T_p\delta_{p0}$, $f^{[1]-}=f_p^{[1]-}\delta_{p0}$, and $f^{[1]+}=f_p^{[1]+}\delta_{p0}$ from the corresponding linear system (A1).

The natural modes of the configuration are obtained by turning off the excitation (i.e., $A^i=0$). The resulting matrix equation possesses a non-trivial solution only if the determinant of the matrix vanishes, i.e.,

$$D_0^i = 2i((\alpha^{[0]i})^2+(\alpha^{[1]i})^2)\sin(k_2^{[1]i}L) - 4\alpha^{[1]i}\alpha^{[0]i}\cos(k_2^{[1]i}L)=0, \tag{14}$$

whose roots take the form of a couple $(k_1^\star,\omega)$, defining the wavenumbers and natural frequencies of the modes of the configuration. For a non-dissipative medium in the plate, $k_1^\star$ and ω are both real. In the case of interest here, the medium filling the plate is dissipative so that $k_1^\star$ is complex. Each mode of the configuration occurs at a wavenumber corresponding to the real part of $k_1^\star$, to which an attenuation corresponding to the imaginary part of $k_1^\star$ is associated.

Figure 2 depicts the real and the imaginary parts of the solution $c^\star(\omega)=\omega/(k_1^\star)$ of Eq. (14) with the mechanical and geometrical characteristics used in Sec. VII A. The solution is normalized by $c^{[0]}$ defining $\bar{c}^\star(\omega)=c^\star(\omega)/c^{[0]}$. The procedure used to solve the dispersion relation latter is based on the one already employed in Ref. 38. For a fixed frequency, the dispersion relation can have more than one root $\bar{c}^\star(\omega)$. They are identified in the following by $\bar{c}^\star_{(n)}(\omega)$, $n\in\mathbb{N}$. An important remark is that the dispersion relation cannot vanish when the incident wave takes the form of a plane wave because $\mathrm{Re}(k_1^\star)\geq k^{[0]}$ for Eq. (14) to be true, whereas $k_1^i\in[-k^{[0]},k^{[0]}]$ in the case of an incident plane (bulk) wave. In the absence of the inclusions, no modes of the configuration can be excited when a plane (bulk) wave strikes the plate.

B. Modal analysis with inclusions

For the plate with periodic inclusions, the problem reduces to the resolution of the linear system (A7). As previously, the natural modes of the configuration are obtained by turning off the excitation, embodied in the vector **VF**. The resulting equation possesses a non-trivial solution only if the determinant of the matrix vanishes:

$$\det[\mathbf{I}-\mathbf{VS}-\mathbf{V}(\mathbf{Q}-\mathbf{P})]=0. \tag{15}$$

A procedure, called the partition method, for solving this equation, is not easy to apply because the off-diagonal elements of the matrix are not small compared to the diagonal elements. Even at low frequency (i.e., $\mathrm{Re}(k^{[1]})R\ll1$), $B_j=B_j^{(2)}\times(k^{[1]}R)^2+O((k^{[1]}R)^2)$ for $j=\{-1,0,1\}$, so that at least three terms should be taken into account.

An iterative scheme can be employed to solve Eq. (A7) and obtain an approximate dispersion relation. This equation is re-written in the form $(1-V_lM_{ll})B_l=V_lF_l+V_l\Sigma_{m\in\mathbb{Z}}M_{lm}B_m(1-\delta_{ml})$, $\forall l\in\mathbb{Z}$. The iterative procedure for solving this linear set of equations is

$$B_l^{(0)}=\frac{V_lF_l}{1-V_lM_{ll}},$$

$$B_l^{(n)}=\frac{V_l\left(F_l+\sum_{m\in\mathbb{Z}}M_{lm}B_m^{(n-1)}(1-\delta_{ml})\right)}{1-V_lM_{ll}}, \tag{16}$$

from which it becomes apparent that the solution $B_l^{(n)}$, to any order of approximation, is expressed as a fraction, the denominator of which $\mathcal{D}_l$ (not depending on the order of approximation) can become small for certain couples (k_{1p},ω) so as to make $B_l^{(n)}$, and (possibly) the field large.

When this happens, a natural mode of the configuration, comprising the inclusions and the plate, is excited, this taking the form of a resonance with respect to $B_l^{(n)}$, i.e., with respect to a plane wave component of the field in the plate relative to the inclusions. As $B_l^{(n)}$ is related to $f_p^{[1]+}$, $f_p^{[1]-}$, T_p, and R_p, the structural resonance manifests itself for the same (k_{1p},ω) in the fields of the plate and in the air.

The approximate dispersion relation

$$\mathcal{D}_l=1-V_l\left[S_0+\sum_{p\in\mathbb{Z}}(Q_{llp}-P_{llp})\right]=0, \tag{17}$$

is the sum of a term linked to the grating embodied in $1-V_lS_0$ with a term linked to the plate embodied in $-V_l\Sigma_{p\in\mathbb{Z}}(Q_{llp}-P_{llp})$, whose expressions are give in Eq. (A4). This can be interpreted as a perturbation of the dispersion relation of the gratings by the presence of the plate.

From Eq. (A6), it is clear that if media $M^{[1]}$ and $M^{[2]}$ have properties that are close (i.e., $\rho^{[1]}\approx\rho^{[2]}$ and $c^{[1]}\approx c^{[2]}$), then the approximate dispersion relation $\mathcal{D}_l=0$ is never satisfied because $V_l\approx0$ (and so $\mathcal{D}_l\approx1$). The contrast between the medium $M^{[1]}$ and $M^{[2]}$ has to be large for the approximate dispersion relation to be satisfied. The zeroth order term of the Schlömilch series S_0 can be re-written[39] in the form $2/d\Sigma_{p\in\mathbb{Z}}1/k_{2p}^{[1]}$ (additional constants are neglected). Introducing this into Eq. (17), together with the expression of Q_{llp} and P_{llp}, gives

$$D_l = 1 - V_l \sum_{p\in\mathbb{Z}} \frac{1}{\dfrac{k_{2p}^{[1]} dD_p}{N_{lp}}} = 0, \quad \forall\, l \in \mathbb{Z}. \tag{18}$$

wherein

$$N_{lp} = 4\cos(k_{2p}^{[1]}(a+b) + 2l\theta_p)((\alpha_p^{[0]})^2 - (\alpha_p^{[1]})^2) - 4\cos(k_{2p}^{[1]}L)((\alpha_p^{[0]})^2 + (\alpha_p^{[1]})^2) + 4i\alpha_p^{[0]}\alpha_p^{[1]}\sin(k_{2p}^{[1]}L).$$

It is then convenient, for the clarity of the explanations, to consider (i) $M^{[1]}$ and $M^{[2]}$ to be non-dissipative media (i.e., medium $M^{[1]}$ and $M^{[2]}$ are perfect fluids) and (ii) the low frequency approximation of R_l (valid when $k^{[1]}R \ll 1$). The latter hypothesis ensures that the V_l reduce to $V_0 = -i\pi/4 \times (K^{[1]} - K^{[2]})/K^{[1]} = -i\pi/4 \times v_0$ and $V_{\pm 1} = -i\pi/4 \times (\rho^{[1]} - \rho^{[2]})/(\rho^{[1]} + \rho^{[2]}) = -i\pi/4 \times v_{\pm 1}$. Equation (18) then reduces to

$$D_l = 1 - \sum_{p\in\mathbb{Z}} \frac{\pi v_l}{\dfrac{4ik_{2p}^{[1]} dD_p}{N_{lp}}} = 0, \quad l = 0, \pm 1. \tag{19}$$

By referring to the work of Cutler,[17] the latter relation should be satisfied for $k_{2p}^{[1]} D_p / N_{lp}$ pure imaginary and close to zero. Because $k^{[1]} < k^{[0]}$ for the authors' application, the first condition (i.e., $k_{2p}^{[1]} D_p / N_{lp}$ pure imaginary) is obtained for $|k_{1p}| \in [k^{[0]}, k^{[1]}]$ (N_{lp} and $k_{2p}^{[1]}$ are real and D_p is pure imaginary in the non-dissipative case). The second condition (i.e., $k_{2p}^{[1]} D_p / N_{lp} = 0$) can then be satisfied only if $D_p = 0$, i.e., when $k_{1p} = k_1^\star$, or if $k_{2p}^{[1]} = 0$, i.e., when $k_{1p} = k^{[1]}$. The latter modes are Rayleigh–Wood modes of the grating (associated with the condition $k_{2p}^{[1]} = 0$) and corresponds to the bounds $|k_{1p}| = k^{[1]}$ defined by the first condition. They can also be hardly excited, especially by a normally incident plane wave. The other modes are neither modes of the plate nor Rayleigh–Wood modes of the grating but modes of the plate together with the grating of inclusions. They satisfy the relation $D_p = 0$ and correspond to evanescent waves in the ambient fluid and propagative wave in the plate. The structure of these modes is close to the one of the modes of the plate without inclusions. They are called *modified plate modes*. To each nth mode $\bar{c}_{(n)}^\star$ of the plate corresponds an infinite number of (n,p) modified plate modes $\bar{c}_{(n,p)}^\star$.

Figure 2 depicts, in the case of the rigid-frame porous (lossy) plate with inclusions, the real and imaginary parts of the solution $c_{(n,p)}^\star(\omega) = \omega / k_{(n,p)}^\star$ normalized by $c^{[0]}$ employing the mechanical characteristics of Sec. VII A, for $d = 10 \times 10^{-3}$ m. $\bar{c}_{(n,p)}^\star(\omega)$ corresponds to the intersection of $\bar{c}_n^\star(\omega)$ as depicted in Sec. IV with $\bar{c}_p(\omega) = \omega/(k_{1p} c^{[0]})$. k_{1p} being real, the associated attenuation corresponds to $\mathrm{Im}(\bar{c}_n^\star(\omega))$ at the frequency $\bar{c}_n^\star(\omega)$ and $\mathrm{Re}(\bar{c}_p(\omega))$ intersect. The modified plate mode can now be excited by an incident plane (bulk) wave, because of the summation over the index p, which allows the existence of evanescent waves in $\Omega_{[0\pm]}$. When a modified plate mode is excited, the structure of the associated waves allow the entrapment of a part of the energy inside the plate and thus to an increase in the absorption.

VI. EVALUATION OF THE REFLECTION, TRANSMISSION, AND ABSORPTION COEFFICIENTS

The developed form of the conservation of energy relation is

$$1 = \mathcal{A} + \mathcal{R} + \mathcal{T}, \tag{20}$$

where $\mathcal{A}$, $\mathcal{R}$, and $\mathcal{T}$ are the absorption, hemispherical reflection, and hemispherical transmission coefficients, respectively. The latter two coefficients are defined by

$$\mathcal{R} = \sum_{p\in\mathbb{Z}} \frac{\mathrm{Re}(k_{2p}^{[0]})\|R_p(\omega)\|^2}{k_2^{[0]i}\|A^i\|^2} = \sum_{p=\bar{p}^-}^{\bar{p}^+} \frac{k_{2p}^{[0]}\|R_p(\omega)\|^2}{k_2^{[0]i}\|A^i\|^2},$$

$$\mathcal{T} = \sum_{p\in\mathbb{Z}} \frac{\mathrm{Re}(k_{2p}^{[0]})\|T_p(\omega)\|^2}{k_2^{[0]i}\|A^i\|} = \sum_{p=\bar{p}^-}^{\bar{p}^+} \frac{k_{2p}^{[0]}\|T_p(\omega)\|^2}{k_2^{[0]i}\|A^i\|}, \tag{21}$$

wherein $\bar{p}$ is such that $(k_1^i + 2\pi(\bar{p} \pm \mp 1)/d)^2 > (k^{[0]})^2 \geq (k_1^i + 2\pi\bar{p}^{\pm}/d)^2$ and the expressions of R_p and T_p are given in Eq. (12), while $\mathcal{A}$ involves non-vanishing surface integrals on Γ_a, Γ_b, and Γ because of viscous effects. Because of the previous argument and because of the complicated shape of $\Omega_{[1]}$ over which some integration should be carried on for the evaluation of the absorption of the porous material itself, $\mathcal{A}$ will be calculated by $\mathcal{A} = 1 - \mathcal{R} - \mathcal{T}$.

VII. NUMERICAL RESULTS AND DISCUSSION

The ambient and saturating fluids are air ($\rho^{[0]} = \rho_f = 1.213$ kg m^{-3}, $c^{[0]} = \sqrt{\gamma P_0/\rho_0}$, with $P_0 = 1.01325 \times 10^5$ Pa and $\gamma = 1.4$, $\eta = 1.839 \times 10^{-5}$ kg m^{-1} s^{-1}). The inclusions are either occupied by melamine-foam (which defines the so-called type 1 sample) or with an acoustically-hard material (for which a Neumann condition prevails on Γ and which defines the so-called type 2 sample). The latter condition is easily encountered for acoustic waves that propagate in rigid-frame porous plate containing inclusions made out of a material whose characteristic impedance is very large compared to that of the porous medium, the latter being very close to the characteristic impedance of the air medium; any elastic materials, such as metals or acrylic plastics, are suitable. Furthermore, the domain Ω_2 does not have to be completely filled with such a material in order for the Neumann boundary condition to prevail; elastic tubes are an example of such inclusions. In practice, this enables the configuration to be lightweight.

The medium $M^{[1]}$ is characterized by $\phi = 0.96$, $\alpha_\infty = 1.07$, $\Lambda = 273 \times 10^{-6}$ m, $\Lambda' = 672 \times 10^{-6}$ m, and $\sigma = 2843$ N s m^{-4}, while the melamine-foam is characterized by $\phi = 0.99$, $\alpha_\infty = 1.001$, $\Lambda = 150 \times 10^{-6}$ m, $\Lambda' = 250 \times 10^{-6}$ m, and $\sigma = 12 \times 10^3$ N s m^{-4}. Melamine foam is more absorptive than the polyurethane foam that occupies the plate, but, on the whole, their mechanical characteristics are rather close. The incident angle is $\theta^i = \pi/2$.

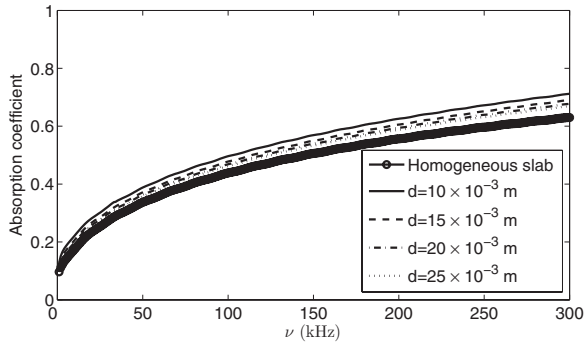


FIG. 3. Absorption coefficients for various center-to center distances and for type 1 samples.

A. Numerical results for an ultrasonic wave incident on a rigid-frame porous plate with a periodic distribution of inclusions

First, an infinite layer 1×10^{-2} m thick and filled with a polymer foam $M^{[1]}$ is considered. Figure 3 depicts the absorption coefficients for type 1, while Fig. 4 depicts the hemispherical reflection, hemispherical transmission, and absorption coefficients for type 2 samples when the radius of the inclusions is equal to $R=2.5\times10^{-3}$ m and the center-to-center distance between inclusion varies from $d=1\times10^{-2}$ m to $d=2.5\times10^{-2}$ m.

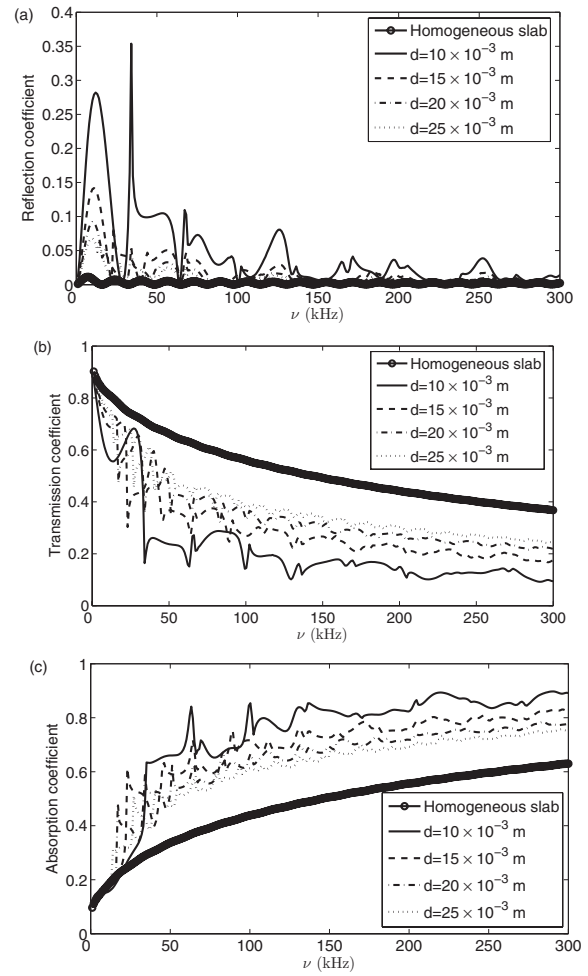


FIG. 4. Hemispherical transmission coefficients (a), hemispherical reflection coefficients (b), and absorption coefficients (c) for various center-to-center distances and for type 2 samples.

For weak contrast type inclusions, no substantial modifications of the hemispherical reflection, transmission, and absorption are noticed. As shown in Sec. VI, when a periodic set of weak contrast inclusions is added to the plate, the dispersion relation cannot vanish so that no modes of the configuration can be excited. The absorption coefficient increases slowly when d decreases while the transmission decreases. The reflection coefficient is practically unmodified by the addition of the periodic set of inclusions. The small increase in the absorption coefficient is due to the fact that a dissipative plate is partially filled with a more dissipative medium, the acoustical properties of the whole configuration being modified in proportion to the ratio of the volume of $M^{[2]}$ to that of $M^{[1]}$.

The results are completely different for large contrast type 2 inclusions for which substantial modifications of the hemispherical reflection, transmission, and absorption are encountered. For frequencies higher than a frequency offset, which depends on the center-to-center distance between inclusions, the hemispherical transmission coefficient decreases substantially, this being associated with an increase in the absorption coefficient. On the contrary, the hemispherical reflection coefficient increases, but less than the decrease in the hemispherical transmission coefficient. The latter seems to be mainly due to the first reflection (in terms of the time domain reflection) on the inclusion when propagative waves are preponderant in the system. Both the hemispherical reflection and transmission coefficient spectra exhibit minima at the same frequencies, at which the absorption coefficient is maximal.

The frequency offset and the maxima of the absorption coefficient (corresponding to minima of the hemispherical reflection and transmission) are linked to, and explained by, the excitation of the modified plate mode. The frequency offset corresponds to the excitation of the first modified plate mode of the configuration. This frequency is also dependent on the thickness of the plate, on the mechanical characteristics of the material that fills the plate, and on the periodicity of the grating. The larger the d is, the smaller the frequency offset is. Below this frequency, no particular modification of the absorption is noticed because waves in the ambient fluid are mainly propagative. Beyond this frequency, evanescent waves in the ambient fluid, associated with modified plate mode excitation, are involved and cause an entrapment of the energy inside the plate, this leading to an increase in the absorption coefficient. This is particularly noticeable for the maxima of the absorption (minima of the hemispherical reflection and transmission) coefficients that are located at the higher modified plate mode natural frequencies. For these frequencies, evanescent waves in the ambient fluid are preponderant and cause an important entrapment of the energy inside the plate.

For a fixed frequency beyond the offset, the increase in the absorption coefficient is dependent (Fig. 3) on the period of the gratings with fixed inclusion radius (large for a small period and small for a large period) and dependent (Fig. 5)

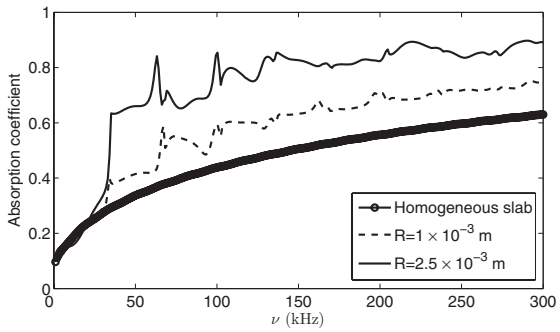


FIG. 5. Comparison of the absorption coefficient for $R=2.5\times10^{-3}$ m and $R=1\times10^{-3}$ m. Type 2 inclusions with $d=10\times10^{-3}$ m.

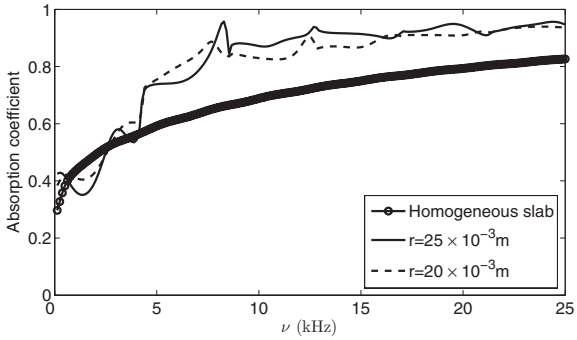


FIG. 7. Absorption coefficients for $R=25\times10^{-3}$ m and $R=20\times10^{-3}$ m. Type 2 inclusions with $d=80\times10^{-3}$ m.

on the radius of the gratings for a fixed period (large for a large radius and small for a small radius). This means that the sought-for increase in the absorption is strongly linked to the determination of the configuration that allows an optimal excitation of modified plate modes. Increasing the period for fixed radius inclusions, or decreasing the radius of the inclusion for a fixed period, both reduce the impact of the inclusions on the global response of the configuration and thus lessen the possibility of modified plate mode excitation. The increase in the absorption coefficient appears to follow the ratio radius over period (r/d).

B. Use of periodic set of high-contrast inclusions for the absorption of audible sound

From the results of Sec. VII A, it can be inferred (i) that the addition of a periodic set of high-contrast inclusions to a rigid-frame porous plate induces an increase in the absorption of the latter for frequencies that are equal or higher than the natural frequency of the first modified plate mode of the configuration and (ii) that the increase in the absorption is all the greater the larger is r/d. The audible frequency range extends from 15–20 Hz to 16–23 kHz. The natural frequency of the first modified plate mode is attained approximately for $\bar{c}_{(1,1)}=(\omega_{(1,1)}d)/(2\pi c^{[0]})\approx0.95$ for normal incident plane waves. The frequency offset is proportional to $1/d$. The lower is the frequency offset, i.e., the lower the resonance frequency of the first modified plate mode, the larger must be the distance that separates adjacent inclusions. The configuration (thickness of the plate, periodicity, and radius) will also be larger for the absorption of audible sound than it is for ultrasound. For example, to increase the absorption coefficient for frequencies that are higher than 4000 Hz, d must be chosen such that $d\approx0.95c^{[0]}/\nu_{(1,1)}\approx81.2\times10^{-3}$ m. $d=80\times10^{-3}$ m is considered hereafter. When compared to the periodicity used in Sec. VII A, d is multiplied by a factor of 8. Two cases are also considered hereafter: $R=25\times10^{-3}$ m and $R=20\times10^{-3}$ m. In practice, the design of acoustical absorbing materials is submitted to constraints in terms of weight and size. The thickness of the plate is thus chosen to be $L=60\times10^{-3}$ m, which corresponds to a radius over the plate thickness ratio $(R/L=25/60\approx0.41)$ larger than for the ultrasonic case $(R/L=2.5/10\approx0.25)$.

Figure 6 depicts the real and imaginary parts of the roots of the dispersion relation in the case of a $L=60\times10^{-3}$ m thick plate filled with $M^{[1]}$ material. The modified plate mode frequencies and associated attenuation are pinpointed in this figure when $d=80\times10^{-3}$ m. The frequency of the first modified plate mode is found to occur at $\nu_{(1,1)}\approx4100$ Hz. Figure 7 depicts the absorption coefficients for type 2 inclusions of radius $R=25\times10^{-3}$ m and $R=20\times10^{-3}$ m. The absorption is larger in case of the larger radius for frequencies higher than $\nu_{(1,1)}\approx4100k$ Hz. The hemispherical reflection and transmission coefficients present minima associated with maxima of the absorption coefficient at the resonance frequencies of the modified plate modes of higher order. The mechanism by the which the absorption is increased is the one described in Sec. VII A (relative to ultrasonic solicitation).

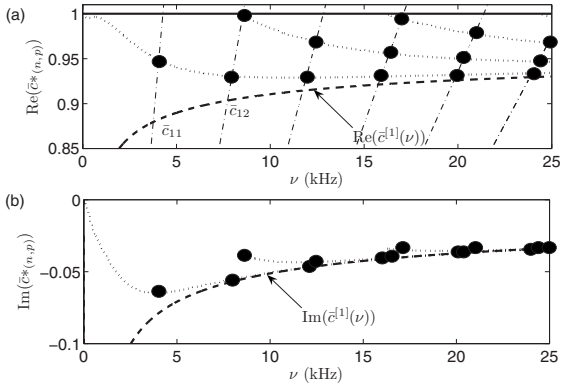


FIG. 6. (a) Modified plate mode and (b) its associated attenuation for a period $d=80\times10^{-3}$ m. The modified plate mode and associated attenuation locations are indicated by dots.

VIII. CONCLUSION

It is shown that high-contrast, periodically-arranged, macroscopic inclusions in a porous plate induce an increase in the absorption coefficient, mainly associated with a decrease in the hemispherical transmission coefficient for frequencies that are higher than a frequency offset. This effect is due to the excitation of the modified plate modes. Expressed differently, the excitation of something similar to a plate mode of the initial macroscopically-homogeneous plate is enabled by the presence of the periodic set of inclusions. The frequency offset of substantial modification of response (with respect to that of the macroscopically-homogeneous plate) corresponds to the frequency of the first modified plate mode of the configuration, while large absorption peaks are

also encountered at the resonance frequencies of the higher order modified plate modes. The type of waves (evanescent in the half spaces above and below the plate, propagative within the plate) associated with these modified plate modes helps to understand why the periodic set of inclusions is necessary for modal excitation and why energy is trapped in the plate at resonance.

The natural frequencies of the modified plate modes were evaluated by referring to the Cutler-modes of gratings through an approximate dispersion relation.

The reflection coefficients were found to be of the same order as those in the absence of inclusions for low-contrast inclusions and to be higher than those in the absence of inclusions for high-contrast inclusions.

Further work should deal with the reduction in the hemispherical reflection coefficient. A possible manner for reducing the latter is by acting on the surface geometry of the plate. A first approach could consist in the addition of a homogenized layer of double porosity.[10]

APPENDIX A: DETERMINATION OF THE UNKNOWNS

1. Application of the continuity conditions across the interfaces Γ_a and Γ_b

Applying successively $\int_{-d/2}^{d/2} \cdot dx_1$ to the continuity of the pressure field and of the normal component of the velocity across Γ_a and Γ_b, Eq. (4), and making use of the orthogonality relation $\int_{-d/2}^{d/2} e^{i(k_{1n}-k_{1l})x_1} dx_1 = d\delta_{nl}$, $\forall (l,n) \in \mathbb{Z}^2$, gives rise to the system of linear equations

$$A^i \delta_{p0} + R_p - f_p^{[1]-} e^{-ik_{2p}^{[1]}a} - f_p^{[1]+} e^{ik_{2p}^{[1]}a} - \sum_{l\in\mathbb{Z}} B_l K_{pl}^{+} e^{ik_{2p}^{[1]}a} = 0, \tag{A1a}$$

$$-A^i \alpha_p^{[0]} \delta_{p0} + R_p \alpha_p^{[0]} + f_p^{[1]-} \alpha_p^{[1]} e^{-ik_{2p}^{[1]}a} - f_p^{[1]+} \alpha_p^{[1]} e^{ik_{2p}^{[1]}a} - \sum_{l\in\mathbb{Z}} B_l K_{pl}^{+} \alpha_p^{[1]} e^{ik_{2p}^{[1]}a} = 0, \tag{A1b}$$

$$T_p - f_p^{[1]-} e^{-ik_{2p}^{[1]}b} - f_p^{[1]+} e^{ik_{2p}^{[1]}b} - \sum_{l\in\mathbb{Z}} B_l K_{pl}^{-} e^{-ik_{2p}^{[1]}b} = 0. \tag{A1c}$$

$$-T_p \alpha_p^{[0]} + f_p^{[1]-} \alpha_p^{[1]} e^{-ik_{2p}^{[1]}b} - f_p^{[1]+} \alpha_p^{[1]} e^{ik_{2p}^{[1]}b} + \sum_{l\in\mathbb{Z}} B_l K_{pl}^{-} \alpha_p^{[1]} e^{-ik_{2p}^{[1]}b} = 0, \tag{A1d}$$

for the resolution of R_p, T_p, and $f^{[1^{\pm}]}$ in terms of B_m, $\forall m \in \mathbb{Z}$ and wherein $\alpha_p^{[i]} = k_{2p}^{[i]}/\rho^{[i]}$ and $i = 0^{\pm}, 1$.

2. Application of the multipole method

Introducing the expressions of $f_p^{[1]-}$ and $f_p^{[1]+}$ in terms of B_l obtained from Eq. (A1) into Eq. (11) leads to an expression of $\tilde{p}^{[1^{\pm}]}(\mathbf{x})$ in the Cartesian coordinate system in terms of B_l, $\forall l \in \mathbb{Z}$. Central to the multipole method are the local field expansion or multipole expansions around each inclusion.[13–15] Because $p(\mathbf{r})$ satisfy a Helmholtz equation inside and outside the cylinder of the unit cell, in the vicinity of the latter (i.e., $\forall \mathbf{r} \le d-R$), $\tilde{p}^{[1]}(\mathbf{r})$ can be written in the form

$$\tilde{p}^{[1]}(\mathbf{r}) = \sum_{l\in\mathbb{Z}} (B_l H_l^{(1)}(k^{[1]}r) + A_l J_l(k^{[1]}r)) e^{il\theta}, \tag{A2}$$

wherein A_l are the coefficients of the locally incident field to the central cylinder, $H_l^{(1)}$ is the lth Hankel function of first kind, and J_l is the lth bessel function. To proceed further, the Cartesian form must be converted to the cylindrical harmonic form by means of $x_1 = r\cos(\theta)$, $x_2 = r\sin(\theta)$, $k_{1p} = k^{[1]}\cos(\theta_p)$, and $k_{2p}^{[1]} = k^{[1]}\sin(\theta_p)$, together with the identity $e^{ikr\cos(\theta-\theta_p)} = \sum_{m\in\mathbb{Z}} i^m J_m(kr) e^{im(\theta-\theta_p)}$:

$$\tilde{p}^{[1]}(\mathbf{r}) = \sum_{l\in\mathbb{Z}} \left[B_l H_l^{(1)}(k^{[1]}r) + \left(\sum_{m\in\mathbb{Z}} M_{lm} B_m + F_l \right) J_l(k^{[1]}r) \right] e^{il\theta}, \tag{A3}$$

with $M_{lm} = S_{l-m} + \sum_{p\in\mathbb{Z}} (Q_{lmp} - P_{lmp})$, $F_l = \sum_{p\in\mathbb{Z}} (J_{lp}^{-} F_p^{[1]-} + J_{lp}^{+} F_p^{[1]+})$, $J_{lp}^{\pm} = i^l e^{\mp il\theta_p}$, and wherein

$$F_p^{[1]\pm} = \left[\pm 2\alpha_p^{[0]} \frac{\alpha_p^{[0]} \mp \alpha_p^{[1]}}{D_p} \right] e^{\mp ik_{2p}^{[1]}b} \delta_{p0} A^i,$$

$$Q_{\mathrm{lmp}} = 4(-i)^{m-l} \frac{(\alpha_p^{[0]})^2 - (\alpha_p^{[1]})^2}{dk_{2p}^{[1]} D_p} \times \cos(k_{2p}^{[1]}(a+b) + (l+m)\theta_p),$$

$$P_{\ln p} = 4(-i)^{m-l} e^{ik_{2p}^{[1]}L} \frac{(\alpha_p^{[0]} - \alpha_p^{[1]})^2}{dk_{2p}^{[1]} D_p} \cos((m-l)\theta_p),$$

$$S_l = \sum_{j=1}^{\infty} H_l^{(1)}(k^{[1]}jd)[e^{ik_1^i jd} + (-1)^l e^{-ik_1^i jd}]. \tag{A4}$$

S_l is the Schlömilch series often referred to as a lattice sum, and D_p as defined Eq. (13).

Identifying Eq. (A2) with Eq. (A3), it follows that

$$A_l = \sum_{m\in\mathbb{Z}} M_{lm} B_m + F_l. \tag{A5}$$

At this point, the two equations (5) and (6) are accounted-for. It is well-known that the coefficients of the scattered field B_l and those of the locally incident field A_l are linked by a matrix relation [derived form the continuity conditions Eqs. (5)] depending on the parameters of the cylinder only $B_l = V_l A_l$, with

$$V_l = \frac{\gamma^{[1]} \dot{J}_l(\chi^{[1]}) J_l(\chi^{[2]}) - \gamma^{[2]} \dot{J}_l(\chi^{[2]}) J_l(\chi^{[1]})}{\gamma^{[2]} \dot{J}_l(\chi^{[2]}) H_l^{(1)}(\chi^{[1]}) - \gamma^{[1]} \dot{H}_l^{(1)}(\chi^{[1]}) J_l(\chi^{[2]})}, \tag{A6}$$

wherein $\gamma^{[j]} = k^{[j]}/\rho^{[j]}$, $\chi^{[j]} = k^{[j]}R$, $j = 0, 1$, and $\dot{Z}_l(x) = dZ_l(x)/dx$. Denoting by $\mathbf{B}$ the infinite column matrix of components B_l, Eq. (A5) together with Eq. (A6) may be written in the matrix form

$$(\mathbf{I} - \mathbf{VM})\mathbf{B} = (\mathbf{I} - \mathbf{VS} - \mathbf{V}(\mathbf{Q} - \mathbf{P}))\mathbf{B} = \mathbf{VF}, \tag{A7}$$

with $\mathbf{F}$ the column matrix of lth element F_l, $\mathbf{I}$ the identity matrix, $\mathbf{V}$ the diagonal matrix of component V_l, and $\mathbf{S}$, $\mathbf{Q}$, and $\mathbf{P}$ three square matrices whose (l,m)th elements are S_{l-m}, $\Sigma_{p\in\mathbb{Z}} Q_{lmp}$, and $\Sigma_{p\in\mathbb{Z}} P_{lmp}$, respectively.

Remark. In the case of a Neumann type boundary condition equation (5), expression Eq. (A6) takes the form $V_l^{\mathcal{N}} = -\dot{J}_l(k^{[1]}R)/\dot{H}_l^{(1)}(k^{[1]}R)$.

APPENDIX B: NUMERICAL RECIPES

The infinite sum $\Sigma_{m\in\mathbb{Z}}$ over the indices of the modal representation of the diffracted field by a cylinder is truncated as $\Sigma_{m=-M}^{M}$ such that[40] $M=\text{int}(\text{Re}(4.05\times(k^{[1]}R)^{1/3}+k^{[1]}R))+10$.

On the other hand, the infinite sum $\Sigma_{p\in\mathbb{Z}}$ over the indices of the k_{1p} is found to depend on the frequency and on the period of the grating. An empirical rule is employed, determined by performing a large number of numerical experiments $\Sigma_{p=-P}^{P}$ such that $P=\text{int}(d/2\pi(3\ \text{Re}(k^{[1]})-k_1^i))+5$. In the latter equations, int(a) represents the integer part of a. Considering the foam plate without dissipation, $k_2^{[1]}$ is the last vertical wave number to become purely imaginary when $|p|$ increases. The previous numerical rule also ensures that $k_{2p}^{[1]}=i\sqrt{2}\|k^{[1]}\|$ (nondissipative case) with an added security term equal 5.

Finally, the infinite sum embedded in S_{m-l} in Eq. (A2) (lattice sum) $\Sigma_{j=1}^{\infty}$ is found to be slowly convergent, particularly in the absence of dissipation, and is found to be strongly dependent on the indices $m-l$. A large literature exists on this problem.[41,42] Here, the fact that the medium $M^{[1]}$ is dissipative greatly simplifies the evaluation of the Schlömilch series. The superscript J in S_{m-l}^{J} identifies the integer over which the sum is performed, i.e., $\Sigma_{j=1}^{J}$. This sum is carried out until the conditions $|\text{Re}((S_{m-l}^{J+1}-S_{m-l}^{J})/S_{m-l}^{J})| \leq 10^{-5}$ and $|\text{Im}((S_{m-l}^{J+1}-S_{m-l}^{J})/S_{m-l}^{J})| \leq 10^{-5}$ are reached.

[1] L. Brekovskikh, *Waves in Layered Media* (Academic, New York, 1960).
[2] U. Ingard, *Notes on Sound Absorption Technology* (Noise Control Foundation, Poughkeepsie, 1994).
[3] O. Tanneau, J. Casimir, and P. Lamary, "Optimization of multilayered panels with poroelastic components for an acoustical transmission objective," J. Acoust. Soc. Am. **120**, 1227–1238 (2006).
[4] M. Biot, "Theory of propagation of elastic waves in fluid-saturated porous solid," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[5] M. Biot, "Mechanics of deformation and acoustic propagation in porous media," J. Appl. Phys. **33**, 1482–1498 (1962).
[6] D. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–402 (1987).
[7] K. Attenborough, "Acoustical characteristics of porous materials," Phys. Rep. **82**, 179–227 (1982).
[8] J.-F. Allard, *Propagation of Sound in Porous Media: Modelling Sound Absorbing Materials* (Chapman & Hall, London, 1993).
[9] L. De Ryck, J.-P. Groby, P. Leclaire, W. Lauriks, A. Wirgin, C. Depollier, and Z. Fellah, "Acoustic wave propagation in a macroscopically inhomogeneous porous medium saturated by a fluid," Appl. Phys. Lett. **90**, 181901 (2007).
[10] X. Olny and C. Boutin, "Acoustic wave propagation in double porosity media," J. Acoust. Soc. Am. **114**, 73–89 (2003).
[11] C. Boutin, P. Royer, and J.-L. Auriault, "Acoustic absorption of porous surfacing with dual porosity," Int. J. Solids Struct. **35**, 4709–4737 (1998).
[12] V. Tournat, V. Pagneux, D. Lafarge, and L. Jaouen, "Multiple scattering of acoustic waves and porous absorbing media," Phys. Rev. E **70**, 026609 (2004).
[13] L. Botten, N. Nicorovici, A. Asatryan, R. McPhedran, C. Poulton, C. de Sterke, and P. Robinson, "Formulation for electromagnetic scattering and propagation through grating stacks of metallic and dielectric cylinders for photonic crystal calculations. Parti method," J. Opt. Soc. Am. A Opt. Image Sci. Vis **17**, 2165–2176 (2005).
[14] D. Felbacq, G. Tayeb, and D. Maystre, "Scattering by a random set of parallel cylinders," J. Opt. Soc. Am. A Opt. Image Sci. Vis **11**, 2526–2538 (1994).
[15] S. Wilcox, L. Botten, R. McPhedran, C. Poulton, and C. de Sterke, "Modeling of defect modes in photonic crystals using the fictitious source superposition method," Phys. Rev. E **71**, 056606 (2005).
[16] R. Wood, "A suspected case of the electrical resonance of minute metal particles for light-waves. A new type of absorption," Proc. Phys. Soc. London **18**,166–182 (1902).
[17] C. Cutler, "Electromagnetic waves guided by corrugated structures," Technical Report No. MM 44-160-218, Bell Telephone Laboratories, 1944.
[18] B. Budiansky, "On the elastic moduli of some heterogeneous materials," J. Mech. Phys. Solids **13**, 223–227 (1965).
[19] Z. Hashin and S. Shtrikman, "A variational approach to the theory of the elastic behaviour of multiphase materials," J. Mech. Phys. Solids **11**, 127–140 (1963).
[20] T. Wu, "The effect of inclusion shape on the elastic moduli of the two-phase material," Int. J. Solids Struct. **1**, 1–8 (1966).
[21] J. Cuomo, J. Ziegler, and J. Woodhall, "A new concept for solar energy thermal conversion," Appl. Phys. Lett. **26**, 557–559 (1975).
[22] C. Horwitz, "Solar-selective globular metal films," J. Opt. Soc. Am. **68**, 1032–1038 (1978).
[23] J. Joannopoulos, R. Meade, and J. Winn, *Photonic Crystals; Molding the Flow of Light* (Princeton University Press, Princeton, 1995).
[24] E. Yablonovitch, "Photonic band-gap structures," J. Opt. Soc. Am. B **10**, 283–295 (1993).
[25] V. Veselago, "The electrodynamics of substances with simultaneous negative value of e and m," Sov. Phys. Usp. **10**, 509–514 (1968).
[26] P.-Y. Bard and A. Wirgin, "Effects of buildings on the duration and amplitude of ground motion in Mexico City," Bull. Seismol. Soc. Am. **86**, 914–920 (1996).
[27] J.-P. Groby and A. Wirgin, "Seismic motion in urban sites consisting of blocks in welded contact with a soft layer overlying a hard half space," Geophys. J. Int. **172**, 725–758 (2008).
[28] A. Khelif, B. Djafari-Rouhani, V. Laude, and M. Solal, "Coupling characteristics of localized phonons in photonic crystal fibers," J. Appl. Phys. **94**, 7944–7946 (2003).
[29] V. Laude, M. Wilm, S. Benchabane, and A. Khelif, "Full band gap for surface acoustic waves in a piezoelectric phononic crystal," Phys. Rev. E **71**, 036607 (2005).
[30] M. Wilm, K. Khelif, S. Ballandras, V. Laude, and B. Djafari-Rouhani, "Out-of-plane propagation of elastic waves in two-dimensional phononic band-gap material," Phys. Rev. E **67**, 065602 (2003).
[31] A. de Bruijn, "Anomalous effects in the sound absorption of periodically uneven surfaces," Acustica **24**, 75–84 (1971).
[32] L. Kelders, J.-F. Allard, and W. Lauriks, "Ultrasonic surface waves above rectangular-groove gratings," J. Acoust. Soc. Am. **103**, 2730–2733 (1998).
[33] B. Sapoval, B. Hebert, and S. Russ, "Experimental study of a fractal acoustical cavity," J. Acoust. Soc. Am. **105**, 2014–2019 (1999).
[34] J.-P. Groby, E. Ogam, and A. Wirgin, "Acoustic response of a periodic distribution of macrosocopic inclusions within a rigid frame porous plate," Waves Random Complex Media **18**, 409–433 (2008).
[35] J.-F. Allard and Y. Champoux, "New empirical equations for sound propagation in rigid frame porous materials," J. Acoust. Soc. Am. **91**, 3346–3353 (1992).
[36] J.-P. Groby, L. De Ryck, P. Leclaire, A. Wirgin, W. Lauriks, R. P. Gilbert, and Y. Xu, "Use of specific Green's functions for solving direct problems involving a heterogeneous rigid frame porous medium slab solicited by the acoustic waves," Math. Methods Appl. Sci. **30**, 91–122 (2007).
[37] J.-P. Groby and C. Tsogka, "A time domain method for modeling viscoacoustic wave propagation," J. Comput. Acoust. **14**, 201–236 (2006).
[38] J.-P. Groby and A. Wirgin, "2D ground motion at a soft viscoelastic layer/hard substratum site in response to SH cylindrical waves radiated by deep and shallow line sources: Numerical results," Geophys. J. Int. **163**, 192–224 (2005).
[39] C. Linton and I. Thompson, "Resonant effects in scattering by periodic arrays," Wave Motion **44**, 165–175 (2007).
[40] P. Barber and S. Hill, *Light Scattering by Particles: Computation Methods*, Advanced Series in Applied Physics 2 (World Scientific, London, 1990), Chaps. 2 and 30.
[41] V. Twersky, "On scattering of waves by the infinite grating of circular cylinders," IRE Trans. Antennas Propag. **10**, 737–765 (1962).
[42] C. Linton, "Schlömilch series that arise in diffraction theory and their efficient computation," J. Phys. A **39**, 3325–3339 (2006).

The forced radiation efficiency of finite size flat panels that are excited by incident sound[a)]

John L. Davy[b)]
School of Applied Sciences, RMIT University, GPO Box 2476V, Melbourne, Victoria 3001, Australia

(Received 3 March 2008; revised 15 May 2009; accepted 15 May 2009)

The radiation efficiency of an infinite flat panel that radiates a plane wave into a half space is equal to the inverse of the cosine of the angle between the direction of propagation of the plane wave and the normal to the panel. The fact that this radiation efficiency tends to infinity as the angle tends to 90° causes problems with simple theories of sound insulation. Sato calculated numerical values of radiation efficiency for a finite size rectangular panel in an infinite baffle whose motion is forced by sound incident at an angle to the normal from the other side. This paper presents a simple two dimensional analytic strip theory, which agrees reasonably well with Sato's numerical calculations for a rectangular panel. This leads to the conclusion that it is mainly the length of the panel in the direction of radiation, rather than its width that is important in determining its radiation efficiency. A low frequency correction is added to the analytic strip theory. The theory is analytically integrated over all angles of incidence, with the appropriate weighting function, to obtain the diffuse sound field forced radiation efficiency of a panel.
 [DOI: 10.1121/1.3158820]

PACS number(s): 43.55.Rg, 43.40.Rj, 43.20.Rz, 43.55.Ti [LMW] Pages: 694–702

I. INTRODUCTION

Below the critical frequency of a panel, the sound transmission through the panel is due mainly to the radiation of sound from the forced bending waves, which are excited by sound incident on the other side. This is because the free bending waves generated by the reflection of the forced bending waves from the edges of the panel have wavelengths that are shorter than the wavelength of sound in air. This means that these free bending waves are very inefficient radiators of sound. At and above the critical frequency of the panel, the free bending waves in the panel have wavelengths that are the same as the forced bending waves excited by sound incident at a frequency dependent angle. Thus to predict the sound insulation of a panel it is necessary to know the radiation efficiency of forced bending waves propagating in the panel.

If a plane wave strikes a panel it forces a bending wave in the panel whose wavelength is greater than or equal to the wavelength of the incident wave in air. Because of this, the forced wave in the panel can radiate efficiently into air on its other side. The radiation efficiency of an infinite panel is equal to the inverse of the cosine of the angle of incidence and transmission [see Sec. IV.7.a of Cremer and Heckl (1988)]. This result obviously cannot be correct for a finite size panel because it goes to infinity at grazing incidence.

Gösele derived the radiation efficiency for an infinite strip in 1953 (Cremer and Heckl, 1988). Sato made numerical calculations in 1973 of the radiation efficiency of a forced wave on a square panel for case where the panel wavelength is longer than the wavelength of sound in air. Sato also numerically calculated the radiation efficiency averaged over all possible directions of sound incidence. Sato's results appear as Fig. 1.3.2, Table B.1, and Fig. B.2 in Rindel, 1975.

Rindel (1975) used Sato's numerical results for radiation efficiency in his theory of sound insulation as a function of angle of incidence. According to Novak (1992), Lindblad produced an approximate formula for the radiation efficiency at high frequencies in 1973 based on Gösele's results. Lindblad derived a simpler approximation in 1985 which can be integrated over all angles of incidence. He also extended the integrated formula to low frequencies.

Rindel (1993a) modified Lindblad's 1973 formula, with constants that are selected to provide good agreement with Sato's tabulated radiation efficiencies. Rindel's formula also extended Lindblad's 1973 formula to low frequencies. This formula of Rindel cannot be integrated by analytic means.

Ljunggren (1991) repeated Sato's numerical calculations using a two dimensional model and obtained agreement "well within 0.5 dB," both as a function of angle of incidence and averaged over all angles of incidence. Novak (1995) performed even more extensive three dimensional calculations than Sato.

The purpose of this paper is to derive an analytic approximation to Sato's numerical results using a simple two dimensional strip model. This analytic approximation has to be simple enough so that it can be integrated by analytic means over all angles of incidence for comparison with Sato's diffuse field results.

[a)]Portions of this work were presented and published in "The radiation efficiency of finite size flat panels," Acoustics 2004, Transportation Noise and Vibration—The New Millennium, Proceedings of the Annual Conference of the Australian Acoustical Society, Gold Coast, 3–5 November 2004.

[b)]Also at CSIRO Materials Science and Engineering, P.O. Box 56, Highett Victoria 3190, Australia. Electronic mails: john.davy@csiro.au and john.davy@rmit.edu.au

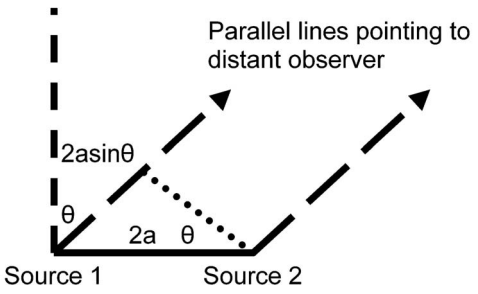


FIG. 1. Two discrete sound sources.

II. DISCRETE AND LINE SOURCES

In this section the radiation of sound from discrete sound sources on a line into a two dimensional plane is considered. This problem is generalized to obtain the radiation from a continuous line source in this section and the radiation from an infinite strip in Sec. III.

Figure 1 shows two point sound sources that are separated by a distance $2a$, which is shown as a solid line. The two sound sources emit pure tones with equal frequency and equal amplitude. An observer at a distance, which is very large compared to the distance $2a$ that separates the sound sources, receives almost the same amplitude sound wave from each source. The lines from the two sound sources to the distant observer, which are shown with long dashes, are almost parallel.

The sound wave from source 1 travels an extra distance $2a \sin\theta$, where θ is the angle between the normal, shown with short dashes, to the line joining the two sound sources and the parallel lines from the two sources to the distant observer. It is also assumed that the phase of source 2 leads the phase of source 1 by 2ψ. Thus, at the distant observer, the phase of the sound from source 2 leads the phase of the sound from source 1 by $2\delta = 2\psi + 2ka \sin\theta$.

Now assume that there are N sources of angular frequency ω in a line of length $2a$. Each source has an amplitude proportional to $1/N$, is a distance $2a/(N-1)$ from the previous source and leads the phase of the previous source by $2\psi/(N-1)$. At the distant observer, the phase of the sound from each source leads the phase of the sound from the previous source by

$$2\delta = \frac{2\psi + 2ka \sin\theta}{N-1}. \tag{1}$$

At time t, the sound pressure at the distant observer is proportional to

$$\frac{1}{N}\sum_{n=1}^{N} \sin[\omega t + 2(n-1)\delta] = \frac{\sin(N\delta)}{N\sin(\delta)} \sin[\omega t + (N-1)\delta]. \tag{2}$$

The above summation is performed using formula 1.341.1 in p. 29 of Gradshteyn and Ryzhik, 1965.

If N is very large

$$N\delta \approx (N-1)\delta = \psi + ka \sin\theta. \tag{3}$$

Thus

$$\delta = \frac{\psi + ka \sin\theta}{(N-1)} \ll 1, \tag{4}$$

and $\sin\delta = \delta$.

Thus the sound pressure at the distant observer is proportional to

$$\frac{\sin(N\delta)}{N\sin(\delta)} \sin[\omega t + (N-1)\delta] = \frac{\sin(\psi + ka\sin\theta)}{\psi + ka\sin\theta} \sin(\omega t + \psi + ka\sin\theta). \tag{5}$$

This large N limit gives the result for a continuous line source of constant source strength over a length of $2a$ and phase difference, which varies linearly by a total amount of 2ψ over the length $2a$ of the continuous line source. The sound pressure amplitude p at a distant observer is proportional to

$$p \propto \frac{\sin(\psi + ka\sin\theta)}{\psi + ka\sin\theta}. \tag{6}$$

If the phase difference ψ is due to a forced bending wave induced by a wave incident at an angle of φ,

$$\psi = -k_b a = -ka\sin\varphi. \tag{7}$$

In this case the sound pressure amplitude p at a distant observer is proportional to

$$p \propto \frac{\sin[ka(\sin\theta - \sin\varphi)]}{ka(\sin\theta - \sin\varphi)}. \tag{8}$$

III. INFINITE STRIPS

Consider an infinite strip of width $2a$ and ask how much power per unit length it radiates from one side when excited by a pure tone wave incident at an angle of φ to the normal to the strip. The wave maxima planes are assumed to be parallel to the two parallel edges of the strip. This is the two dimensional problem considered in Sec. II. Squaring the amplitude given by Eq. (8) produces the power that can be summed by integrating over all angles of radiation θ from $-\pi/2$ rad to $\pi/2$ rad. Make the following approximation

$$\sin\theta - \sin\varphi = 2\sin\left(\frac{\theta-\varphi}{2}\right)\cos\left(\frac{\theta+\varphi}{2}\right) \approx (\theta-\varphi)\cos\varphi \quad \text{for } |\theta-\varphi| \ll 1. \tag{9}$$

If the approximation given by Eq. (9) is inserted into the square of Eq. (8), the resulting expression is only significantly different from zero when φ is near θ. Because θ is between $-\pi/2$ and $\pi/2$, it is possible to approximate by extending the limits of integration to $-\infty$ to ∞. These approximations will give the infinite panel result. Later in this paper, the range of validity of these approximations is examined and other approximations are obtained for use outside the range of validity of these approximations. A low frequency correction will also be introduced in the Sec. IV. With the above approximations, the total radiated sound power per unit length of strip is proportional to

$$\int_{-\infty}^{\infty} \frac{\sin^2[ka(\theta-\varphi)\cos\varphi]}{[ka(\theta-\varphi)\cos\varphi]^2} d\theta = \int_{-\infty}^{\infty} \frac{\sin^2(ka\theta\cos\varphi)}{(ka\theta\cos\varphi)^2} d\theta$$

$$= \frac{\pi}{ka\cos\varphi}, \tag{10}$$

using integral 3.821.9 in p. 446 of Gradshteyn and Ryzhik, 1965.

This is the same $1/\cos\varphi$ variability as in the case of the infinite panel since for the infinite panel, the transmitted angle θ is equal to the incident angle φ. Equation (10) is only proportional to the radiation efficiency of the infinite strip. Since the radiation efficiency of an infinite strip must equal the $1/\cos\varphi$ radiation efficiency of an infinite panel if ka is large enough, Eq. (10) must be multiplied by ka/π to obtain the absolute value of the radiation efficiency. Thus the radiation efficiency of an infinite strip for an incident angle of φ is

$$\sigma(\varphi) = \frac{1}{\cos\varphi} = \frac{1}{g}, \tag{11}$$

where $g=\cos\varphi$. This result is obtained by Gösele, Lyon and Maidanik (1962), and Maidanik (1962).

The range of validity of Eq. (11) is now investigated. The maximum value of the integrand on the left side of Eq. (10) is 1 when $\theta=\varphi$. Thus replace this integrand in Eq. (10) with a function which is equal to 1 when

$$|\theta-\varphi| \le \frac{\pi}{2ka\cos\varphi}, \tag{12}$$

and is zero elsewhere. This function gives the same value for the integral. For this replacement function, the change in the limits of integration is only valid if the nonzero part of the replacement function lies between $-\pi/2$ and $\pi/2$. This means that

$$\frac{\pi}{2} - |\varphi| \ge \frac{\pi}{2ka\cos\varphi}. \tag{13}$$

For $|\varphi|$ close to $\pi/2$, the left hand side of Eq. (13) is approximately $\cos\varphi$. Thus Eq. (13) becomes

$$\cos\varphi \ge \frac{\pi}{2ka\cos\varphi}, \tag{14}$$

or

$$\cos\varphi \ge \sqrt{\frac{\pi}{2ka}}, \tag{15}$$

or

$$|\varphi| \le \arccos\sqrt{\frac{\pi}{2ka}}. \tag{16}$$

Thus Eqs. (10) and (11) are only valid in the range given by Eq. (16). At the two angles of incidence φ_l given by the equal sign in Eq. (16), the total radiated sound power per unit length of the strip is proportional to

$$\frac{\pi}{ka\cos\varphi_l} = \frac{\pi}{ka}\sqrt{\frac{2ka}{\pi}} = \sqrt{\frac{2\pi}{ka}}. \tag{17}$$

For later use, Eq. (15) is rewritten as

$$g = \cos\varphi \ge \begin{cases} g_l = \sqrt{\dfrac{\pi}{2ka}} & \text{if } g_l \le 1, \\ 1 & \text{if } g_l > 1, \end{cases} \tag{18}$$

where g_l is defined by the second equality in Eq. (18).

In practice, an empirical correction factor w is used in Eq. (18) for the range of Eqs. (10) and (11).

$$g = \cos\varphi \ge f = \begin{cases} wg_l = w\sqrt{\dfrac{\pi}{2ka}} & \text{if } wg_l \le 1, \\ 1 & \text{if } wg_l > 1, \end{cases} \tag{19}$$

where f is defined by the second equality in Eq. (19). The empirical correction factor w will be determined by comparison with Sato's numerical calculations for a square panel.

Since the maximum value of the integrand on the left side of Eq. (10) is 1, the maximum value of the integral before extending the limits is $\pi/2\text{-}(-\pi/2)=\pi$. Also $\cos\varphi$ is in the range from 0 to 1 for all values of φ in the range from $-\pi/2$ to $\pi/2$. This gives

$$\frac{\pi}{ka} \le \frac{\pi}{ka\cos\varphi} \le \pi. \tag{20}$$

This means that the approximations can only be valid if $ka \ge 1$.

It is also possible to approximate the integral if $|\varphi| = \pi/2$. Because of symmetry in the equations, only the case $\varphi=\pi/2$ is considered. Now

$$\sin(\theta) - \sin(\varphi) = \cos\left(\frac{\pi}{2} - \theta\right) - 1. \tag{21}$$

If $\pi/2-\theta$ is small, Eq. (21) becomes

$$1 - \frac{1}{2}\left(\frac{\pi}{2} - \theta\right)^2 - 1 = -\frac{1}{2}\left(\frac{\pi}{2} - \theta\right)^2. \tag{22}$$

Put $y=\pi/2-\theta$, then

$$ka[\sin(\theta) - \sin(\varphi)] = -kay^2/2. \tag{23}$$

The integral in Eq. (10) becomes

$$\int_0^{\infty} \frac{\sin^2(kay^2/2)}{(kay^2/2)^2} dy. \tag{24}$$

The $\theta=\pi/2$ limit has become $y=0$. The $\theta=-\pi/2$ limit has become $y=\pi$ and been extended to $y=\infty$. This extension will be compensated by the introduction of an empirical offset correction β in Eq. (27) and the introduction of a low frequency correction in Sec. IV.

Using integral number 3.852.3 in p. 464 of Gradshteyn and Ryzhik (1965), Eq. (24) becomes

$$\left(\frac{2}{ka}\right)^2 \frac{2\sqrt{\pi}}{3}\left(\frac{ka}{2}\right)^{3/2} = \frac{2}{3}\sqrt{\frac{2\pi}{ka}}. \tag{25}$$

It should be noted that this result is 2/3 of the value derived in Eq. (17) for when

$$\cos\varphi_l = \sqrt{\frac{\pi}{2ka}}. \tag{26}$$

Like Eq. (10), Eq. (25) must be multiplied by ka/π to obtain the absolute value of the radiation efficiency. The radiation efficiency of an infinite strip for a grazing incident angle φ for which $|\varphi| = \pi/2$ rad is

$$\sigma\left(\frac{\pi}{2}\right) = \frac{1}{h} = \frac{2}{3}\frac{1}{g_l} - \beta = \frac{2}{3}\sqrt{\frac{2ka}{\pi}} - \beta, \tag{27}$$

where

$$h = \frac{1}{\frac{2}{3g_l} - \beta} = \frac{1}{\frac{2}{3}\sqrt{\frac{2ka}{\pi}} - \beta}, \tag{28}$$

and β is an empirical offset correction, which is determined by comparison with Sato's numerical calculations for a square panel. It is thought that β is needed because the theory is developed for an infinite strip but is applied to the case of a square panel. β accounts for the fact that the radiation efficiency of an infinite strip is slightly greater than that of a square panel at grazing angles of excitation.

This result with $\beta=0$ is derived by Gösele. Lyon and Maidanik (1962) also derived this result with $\beta=0$, but a factor of 2 appears to be missing from their equation. Cremer and Heckl (1988) insert the factor of 2 when quoting Lyon and Maidanik's (1962) result. Maidanik (1962) derived this result with $\beta=0$ but then appeared to include an extra factor of $2\times1.06=2.12$ in his final result for the radiation efficiency at grazing angles of incidence. The source of the 1.06 is explained by Eqs. (31) and (32) of this paper. This overestimate of Maidanik (1962) in his final result for the radiation efficiency at grazing angles of incidence is confirmed by Novak's (1995) three dimensional numerical calculations on a square panels of side length $2a$ for $ka=16$ and $ka=64$. For a square panel with sides of length $2a$, Maidanik's (1962) final result for the radiation efficiency at grazing angles of incidence is

$$\langle\sigma\rangle = \sqrt{\frac{2a}{\lambda}} + \sqrt{\frac{2a}{\lambda}} = 2\sqrt{\frac{2a}{\lambda}}. \tag{29}$$

For a square panel with sides of length $2a$, Cremer and Heckl (1988) quoted Maidanik's (1962) final result for the radiation efficiency at grazing angles of incidence as

$$\langle\sigma\rangle = 0.45\sqrt{\frac{U}{\lambda}} = 0.45\sqrt{\frac{8a}{\lambda}} = 0.9\sqrt{\frac{2a}{\lambda}}, \tag{30}$$

where U is the perimeter of the panel. The ratio of Eq. (29) to Eq. (30) of $2/0.9=2.22$ is close to the extra factor of 2.12, which appears to be included in Eq. (29). Leppington *et al.* (1982) also showed that Maidanik's (1962) final result for the radiation efficiency at grazing angles of incidence for rectangular panels with aspect ratios between 0.3 and 1 overestimates "by a factor of about 2." It should be noted that Maidanik (1962) only claimed that his "values should be accurate at least to an order of magnitude."

If $\beta=0$,

$$h = \frac{3}{2}\sqrt{\frac{\pi}{2ka}} = \frac{3}{2\sqrt{2}}\sqrt{\frac{\lambda}{2a}} = 1.06\sqrt{\frac{\lambda}{2a}}. \tag{31}$$

Because 1.06 is approximately equal to 1, Maidanik (1962), Lindblad, and Novak (1992) used the following approximation:

$$h = \sqrt{\frac{\lambda}{2a}} \text{ and } \sigma\left(\frac{\pi}{2}\right) = \frac{1}{h} = \sqrt{\frac{2a}{\lambda}}. \tag{32}$$

The range of Eq. (11) is extended beyond that given by Eq. (19) by combining the infinite panel result of $1/\cos\varphi$ [see Eq. (11)] with the infinite strip result at a grazing angle of incidence of Eq. (27). Linear interpolation in $g=\cos\varphi$ in the denominator of Eq. (11) from f when $g=\cos\varphi=f$ to h when $g=\cos\varphi=0$ is used. This gives

$$\sigma(g) = \begin{cases} \dfrac{1}{g} & \text{if } 1 \geq |g| \geq f, \\ \dfrac{1}{h - \alpha g} & \text{if } f > |g| \geq 0, \end{cases} \tag{33}$$

where the interpolation slope α is chosen so that the two parts of Eq. (33) are equal when $g=\cos\varphi=f$. This gives

$$\alpha = \frac{h}{f} - 1. \tag{34}$$

IV. FINITE SIZE PANELS

The approximation of extending the limits of integration from $-\pi/2$ and $\pi/2$ to $-\infty$ and ∞ in Eq. (10) and from π to ∞ in Eq. (24) means that Eq. (33) will overestimate the radiation efficiency at low frequencies where ka is small. To extend the results to values of ka less than 1, assume a finite size panel of area S. Since only the power that is radiated is of interest, only the real part of the fluid wave impedance Z_{wf} is considered. For a symmetrically pulsating sphere of radius r, the real part of the normalized fluid wave impedance for $kr \ll 1$ is k^2r^2. By symmetry this result also applies to a pulsating hemisphere whose center is on an infinite rigid plane. For sources whose size is small compared to the wavelength of sound, it is expected that the real parts of their normalized fluid wave impedances will depend only on their areas. Thus the result for the pulsating hemisphere also applies to a panel set in an infinite rigid plane baffle provided the area of the panel is equal to the surface area of the hemisphere. Thus $2\pi r^2 = S$ where S is the area of the panel. For a square panel of side length $2a$ the area of the panel is $S=4a^2$. The radiation efficiency of the panel is

$$\sigma = \frac{\mathrm{Re}(Z_{wf})}{\rho_0 c} = k^2r^2 = \frac{k^2S}{2\pi} = \frac{1}{q} = \frac{2}{\pi}k^2a^2, \tag{35}$$

where the inverse of the radiation efficiency of the panel is

$$q = \frac{2\pi}{k^2 S} = \frac{\pi}{2k^2 a^2}. \tag{36}$$

Equations (35) and (36) depend on the assumption that all areas of the square panel are vibrating in phase with uniform amplitude. For a panel that is excited by incident airborne sound and whose dimensions are small compared to the wavelength of sound in air, this is a reasonable assumption. For non-square panels, the characteristic length to use in Eqs. (35) and (36) is $2a=\sqrt{S}$, where S is the area of the panel.

Equations (33) and (35) are combined by inverting, raising to the power of n, adding, taking the nth root and inverting. Lindblad, Novak (1992), and Rindel (1993a) obtain their versions of Eq. (33) by using this method to combine the two different incident angle ranges. Rindel (1993a) also used it to combine his version of Eq. (33) with Eq. (35). This combining method gives the radiation efficiency as

$$\sigma(g) = \begin{cases} \dfrac{1}{\sqrt[n]{g^n + q^n}} & \text{if } 1 \geq |g| \geq f, \\ \dfrac{1}{\sqrt[n]{[h-\alpha g]^n + q^n}} & \text{if } f > |g| \geq 0. \end{cases} \tag{37}$$

The empirical combining power n will be determined by comparison with Sato's numerical calculations for a square panel.

V. AVERAGE OVER AZIMUTHAL ANGLE

For a specific azimuthal incidence direction, $2a$ should be set equal to a characteristic length of the panel in the azimuthal incidence direction. For averages over all azimuthal incidence directions, Ljunggren (1991) suggested that $2a$ should be set equal to the mean free path $\pi S/U$, where S is the area and U is the perimeter of the panel. Ljunggren (1991) also showed that the root mean square length across a rectangle is equal to $\sqrt{2S/\pi}$ and points out that this "is in most cases, in practice, fairly close to the mean free path." Novak (1995) confirmed the usefulness of Ljunggren's (1991) suggestion by showing that the diffuse field forced radiation efficiencies of a square panel and a rectangular panel are almost identical for the same value of $k\pi S/U$.

The problem with $\pi S/U$ and $\sqrt{2S/\pi}$ is that they are equal to $(\pi/4)2a=0.785\times 2a$ and $\sqrt{(2/\pi)}2a=0.798\times 2a$, respectively, rather than $2a$, for a square panel with side length $2a$. All that Novak's (1995) result actually shows is that the characteristic length should be proportional to S/U. Rindel (1993a, 1993b) followed Thomasson's (1982) proposal of a characteristic length of $4S/U$. This is consistent with Novak's (1995) result and is equal to $2a$ for a square panel with side length $2a$. It is also equal to the diameter for a circular panel.

As shown in Eq. (30), Cremer and Heckl (1988) used a characteristic length of $U/4$ when quoting Maidanik's (1962) final result for the radiation efficiency at grazing angles of incidence. This is equal to $2a$ for a square panel with side length $2a$. Sewell (1970) used a characteristic length of $\sqrt{S}$. This is also equal to $2a$ for a square panel with side length $2a$ and is the characteristic length that is used in Eqs. (35) and (36). One of Maidanik's (1962) equations can also be interpreted as using a characteristic length of $\sqrt{S}$ at grazing incidence. However, Sewell (1970) also used a correction offset for non-square rectangular panels. This shows that $\sqrt{S}$ is not completely satisfactory as a characteristic length for non-square panels. Thus, this paper recommends the use of a characteristic length of $2a=4S/U$ for non-square panels, except in Eqs. (35) and (36), where the characteristic length $2a=\sqrt{S}$ is used.

Leppington *et al.* (1982) showed that the radiation efficiency, at grazing incidence averaged over all azimuthal angles of incidence, of a rectangular panel with a shorter side length of $2L$ and a longer side length of $2W$ is

$$\sigma = H(x)\sqrt{2kL} \quad \text{where } x = \frac{2L}{2W}. \tag{38}$$

Leppington *et al.* (1982) gave a formula for $H(x)$, which involves a fairly complicated integral and a graph of $H(x)$ for x between 0 and 1. They said that the approximation $H_1(x)=0.5-0.15x$ differs from $H(x)$ by less than 4%. For the same rectangular panel and a characteristic length of $2a=4S/U=4L/(1+x)$, Eq. (27), with $\beta=0$, can be written as

$$\sigma\left(\frac{\pi}{2}\right) = H_2(x)\sqrt{2kL} \text{ where } H_2(x) = \frac{2}{3}\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{1+x}}$$

$$\text{and } x = \frac{2L}{2W}. \tag{39}$$

$H_2(x)$ differs from $H(x)$ by less than about 5% and from $H_1(x)$ by less than about 6%. Thus the work of Leppington *et al.* (1982) also justifies the use of a characteristic length of $2a=4S/U$.

VI. AVERAGE OVER ANGLE OF INCIDENCE

The radiation efficiency averaged over all angles of incidence φ and all azimuthal angles is

$$\langle\sigma\rangle = \int_0^{\pi/2} \sigma(\varphi)\sin\varphi d\varphi. \tag{40}$$

The $\sin\varphi$ occurs in the integral because there is more solid angle for sound to be incident from the closer φ is to $\pi/2$. To evaluate this integral, assume the constant value of the characteristic $2a=4S/U$ given in Sec. V.

Differentiating $g=\cos\varphi$ gives $dg=-\sin\varphi d\varphi$. Equation (40) becomes

$$\langle\sigma\rangle = \int_f^1 \frac{dg}{\sqrt[n]{g^n + q^n}} + \int_0^f \frac{dg}{\sqrt[n]{(h-\alpha g)^n + q^n}}. \tag{41}$$

If $\alpha\neq 0$, then put $x=h-\alpha g$. Differentiating gives $dx=-\alpha dg$ and Eq. (41) becomes

$$\langle\sigma\rangle = \int_f^1 \frac{dg}{\sqrt[n]{g^n + q^n}} + \frac{1}{\alpha}\int_f^h \frac{dx}{\sqrt[n]{x^n + q^n}}. \tag{42}$$

If $n=2$, integral number 2.271.4 in p. 86 of Gradshteyn and Ryzhik, 1965 gives

$$\langle\sigma\rangle = \ln\left(\frac{1+\sqrt{1+q^2}}{f+\sqrt{f^2+q^2}}\right) + \frac{1}{\alpha}\ln\left(\frac{h+\sqrt{h^2+q^2}}{f+\sqrt{f^2+q^2}}\right). \quad (43)$$

If $n=1$, integral number 2.111.1 in p. 58 of Gradshteyn and Ryzhik, 1965 gives

$$\langle\sigma\rangle = \ln\left(\frac{1+q}{f+q}\right) + \frac{1}{\alpha}\ln\left(\frac{h+q}{f+q}\right). \quad (44)$$

Another possible way of combining the infinite panel result with the infinite strip result for grazing incidence, which was introduced by Lindblad in 1985, is obtained by setting $\alpha=0$ and

$$f = \begin{cases} h & \text{if } h \le 1, \\ 1 & \text{if } h > 1. \end{cases} \quad (45)$$

This paper will describe this method as Lindblad's simpler method.

If $\alpha=0$, Eqs. (43) and (44) must be replaced with the following two equations, respectively. For $n=2$,

$$\langle\sigma\rangle = \ln\left(\frac{1+\sqrt{1+q^2}}{f+\sqrt{f^2+q^2}}\right) + \frac{f}{\sqrt{h^2+q^2}}. \quad (46)$$

For $n=1$,

$$\langle\sigma\rangle = \ln\left(\frac{1+q}{f+q}\right) + \frac{f}{h+q}. \quad (47)$$

Lindblad also combined the infinite panel result with the infinite strip result at grazing incidence using the same combining method as used to obtain Eq. (37) in this paper. That is, he inverted, raised to the power n, summed, took the nth root, and inverted again. Lindblad used $n=4$. This paper will describe this method as Lindblad's more complicated method. Novak (1992) used this method with $n=10$. The disadvantage of using this method with $n=4$ or 10 is that it is not easy to analytically evaluate the integral of Eq. (40). Lindblad only applied the low frequency correction to the integrated version of his simpler approximation. Lindblad and Novak (1992) did not combine their results with the low frequency result of eq. (35). In this paper their results are combined with the low frequency result using $n=2$.

Rindel's approximation (Rindel, 1993a) is the only other formula to produce a peak at other than grazing incidence or at normal incidence. It is also the only other formula to include the low frequency correction q. It combines the low angle of incidence result with both the grazing incidence result and the low frequency result using the $n=4$ method, but modifies the low angle of incidence result and the grazing incidence result so that the maximum occurs at $g=1.1g_l$. This compares with the method developed in this paper which has the maximum at $g=1.3g_l$. The observation that Rindel's maximum occurs at a value of g different to g_l is one of the reasons for the introduction of the empirical correction factor w in this paper. Unfortunately Rindel's formula is too complicated to be easily analytically integrated. This is why Rindel (1993b) also developed a separate approximation to the diffuse field radiation efficiency.

For comparison with some other results, ignore the low frequency result by setting $q=0$, set $w=1$ and $\beta=0$, and assuming that $f \leqslant 1$. These assumptions give $h/f=3/2$ and Eqs. (43) and (44) both reduce to

$$\langle\sigma\rangle = \tfrac{1}{2}(\ln 2ka + 0.477). \quad (48)$$

Rindel (1993b) gave the following diffuse field radiation efficiency approximation which he said is useful for $ka>0.5$.

$$\langle\sigma\rangle = \tfrac{1}{2}(\ln 2ka + 0.2). \quad (49)$$

Sewell's (1970) work can be interpreted as producing a similar formula with a low frequency correction. This is because the diffuse field sound transmission coefficient for a single isotropic panel is obtained by multiplying the panel's normal incidence transmission coefficient by twice Eq. (40).

$$\langle\sigma\rangle = \frac{1}{2}\left(\ln 2ka + 0.160 + \frac{1}{16\pi k^2 a^2}\right). \quad (50)$$

Setting $ka=1$ in the third term in the brackets in Eq. (63) gives

$$\langle\sigma\rangle = \tfrac{1}{2}(\ln 2ka + 0.180). \quad (51)$$

Sewell (1970) also gave a correction for non-square rectangular panels, which reduce the diffuse field radiation efficiency.

Adopting Lindblad's assumptions of $q=0$, $\alpha=0$, f given by Eq. (45), h given by Eq. (32), and $f \leqslant 1$ gives $f=h$ and Eqs. (46) and (47) both reduce to Lindblad's equation for the diffuse field radiation efficiency.

$$\langle\sigma\rangle = \ln\left(\sqrt{\frac{ka}{\pi}}\right) + 1 = \frac{1}{2}(\ln 2ka + 0.162). \quad (52)$$

If Sewell's (1970) low frequency correction term is ignored, his Eq. (50) is almost the same as Lindblad's Eq. (52) and very similar to Rindel's Eq. (49). Comparison of these three equations with Eq. (48) suggests that Eq. (48) slightly overestimates the diffuse field radiation efficiency. This is one of the reasons for applying the offset correction term β to the grazing incidence radiation efficiency.

Equation (37) of this paper can also be used to predict the sound radiation of free bending waves in a panel if the angular frequency ω is greater than or equal to the angular critical frequency of the panel ω_c. Because the free bending wavelength of the panel is greater than or equal to the wavelength of sound in air λ and correspondingly the free bending wave number of the panel k_b is less than or equal to the wave number of sound in air k, $\sin\theta$ can be calculated from $k_b = k\sin\theta$. This enables $g=\cos\theta$ to be calculated.

$$g = \cos\theta = \sqrt{1-\sin^2\theta} = \sqrt{1-\left(\frac{k_b}{k}\right)^2} = \sqrt{1-\frac{\omega_c}{\omega}}. \quad (53)$$

Note that Equation (2.39c) in Maidanik (1962) mistakenly interchanges the critical frequency and the frequency. This mistake is not corrected in the errata.

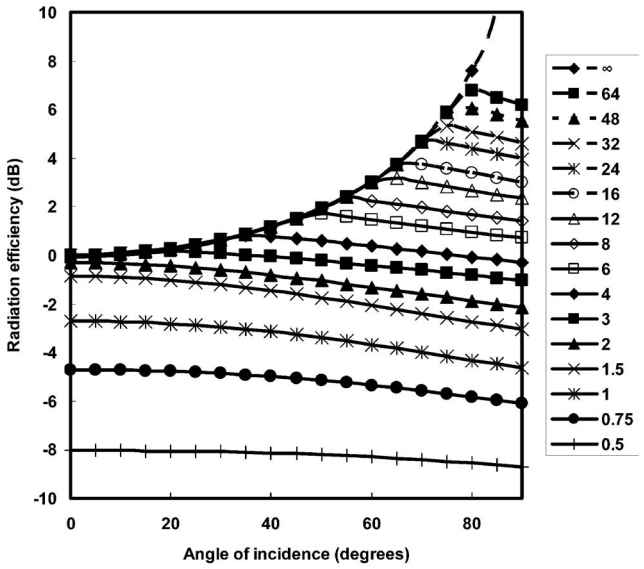


FIG. 2. The theoretical forced radiation efficiency of a square panel, with side length of $2a$, as a function of the angle of incidence for a range of values of ka. These theoretical results have been calculated using Eqs. (19), (28), (34), (36), and (37).

VII. COMPARISON WITH CALCULATION AND EXPERIMENT

Davy (2004) used $n=1$, $w=1$, and $\beta=0$ and in the associated conference presentation gave results for the case $n=2$, $w=1$, and $\beta=0$. This paper uses $n=2$, $w=1.3$, and $\beta=0.124$ because these values produce the best agreement with Sato's numerically calculated results. Figure 2 shows the calculated values of the radiation efficiency in decibel as a function of the angle of incidence from 0° to 90° for values of ka ranging from 0.5 to ∞.

Sato produced a table of numerically calculated values of the forced radiation efficiency in dB to one decimal place for a square panel with side length $2a$, at 15° increments of the angle of incidence from 0° to 90° to the normal to the panel for 15 values of ka from 0.5 to 64 inclusive. Table I shows the difference in decibel between Eq. (37) and Sato's numerical results. The biggest errors occur because of ripple in the change from the high frequency region to the low frequency region around $ka=2$. This is why most other authors did not extended their approximations to low frequencies. Table II shows the effect on the differences at an angle of incidence of 90°, if β is changed from 0.124 to 0.

TABLE I. The differences (dB) between the radiation efficiency given by Eq. (37) and Sato's numerically calculated radiation efficiency.

ka	0°	15°	30°	45°	60°	75°	90°
0.5	0.3	0.3	0.2	0.2	0.2	0.1	−0.1
0.75	0.3	0.3	0.3	0.3	0.3	0.0	−0.4
1	0.3	0.3	0.3	0.3	0.2	0.0	−0.4
1.5	−0.3	−0.2	−0.1	0.1	0.1	0.0	−0.4
2	−0.7	−0.6	−0.4	−0.1	0.0	0.1	−0.1
3	0.0	−0.1	−0.4	−0.5	−0.2	0.0	−0.1
4	0.3	0.1	−0.1	−0.3	−0.2	0.0	0.0
6	0.0	−0.1	0.0	−0.1	−0.2	0.0	0.0
8	−0.1	0.0	0.1	−0.1	−0.1	0.0	0.0
12	0.0	0.0	0.0	0.1	−0.1	−0.1	0.1
16	0.0	0.0	0.0	0.0	−0.2	−0.1	0.1
24	0.0	0.0	0.0	0.0	0.1	−0.1	0.0
32	0.0	0.0	0.0	0.0	0.0	−0.1	0.0
48	0.0	0.0	0.0	0.0	0.0	−0.1	0.0
64	0.0	0.0	0.0	0.0	0.0	−0.2	0.0

TABLE II. The differences (dB) between the radiation efficiency given by Eq. (37) and Sato's numerically calculated radiation efficiency for an angle of incidence of 90° when $\beta=0.124$ and $\beta=0$.

ka	$\beta=0.124$	$\beta=0$
0.5	−0.1	0.3
0.75	−0.4	0.2
1	−0.4	0.3
1.5	−0.4	0.4
2	−0.1	0.6
3	−0.1	0.5
4	0.0	0.5
6	0.0	0.4
8	0.0	0.4
12	0.1	0.4
16	0.1	0.3
24	0.0	0.3
32	0.0	0.2
48	0.0	0.2
64	0.0	0.1

Sato also graphed numerically calculated values of the forced radiation efficiency in decibel for a square panel with side length $2a$, at 5° increments of the angle of incidence from 0° to 90° to the normal to the panel for six values of ka from 2 to 64 inclusive. The extra values on this graph are read from Sato's graph and added to his tabulated values. This enlarged set of values is used to compare the results of Eq. (37) (D), Rindel's (1993a) approximation (R1), Lindblad's more complicated approximation (L1), Lindblad's simpler approximation (L2), and Novak's (1992) approximation (N) to Sato's numerically calculated results. The mean, the standard deviation, the maximum, and the minimum differences from Sato's numerically calculated results are shown in Table III. Equation (37) performs slightly better than the other approximation formulae.

Figure 3 shows the calculated values of the diffuse field forced radiation efficiency in decibel as a function ka from 0.5 to 64. Sato also tabulated numerically calculated values of the diffuse field forced radiation efficiency in decibel to two decimal places, for a square panel with side length $2a$, for 15 values of ka from 0.5 to 64 inclusive. Table IV shows the differences in decibel between Eq. (43) (D), Lindblad's

TABLE III. The mean, the standard deviation, the maximum, and the minimum differences (dB) of various approximations from Sato's numerically calculated radiation efficiency results.

	D	R1	L1	L2	N
Mean	0.0	0.0	−0.2	0.0	0.0
St. dev.	0.2	0.4	0.5	0.4	0.5
Maximum	0.3	0.9	0.8	0.8	0.8
Minimum	−0.7	−1.5	−1.9	−1.5	−1.6

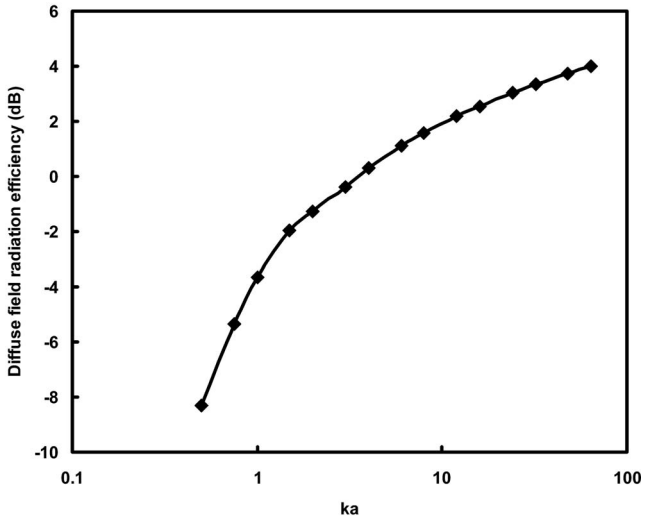


FIG. 3. The theoretical diffuse field forced radiation efficiency of a square panel, with side length of $2a$, as a function of ka. These theoretical results have been calculated using Eqs. (19), (28), (34), (36), and (43).

diffuse field result from the integrated version of his simplified approximation (L2), Rindel's (1993b) diffuse field radiation efficiency approximation (R2), and Sewell's (1970) formula (S). The mean, the standard deviation, the maximum, and the minimum differences from Sato's numerically calculated results are also shown in Table IV. Rindel (1993b) said that his approximation is useful for $ka > 0.5$. Sewell (1970) also gave a correction term for non-square rectangular panels. The integral of Lindblad's simpler approximation formula performs slightly better than the other formulae. If Lindblad's simpler approximation formula is integrated without the low frequency correction to obtain Eq. (52) and the low frequency correction of Eq. (35) is combined with it using the $n=2$ method, the result agrees with the L2 results in Table III to better than three decimal places.

TABLE IV. The differences (dB) between various diffuse field radiation efficiency approximations and Sato's numerically calculated diffuse field radiation efficiency.

ka	D	L2	R2	S
0.5	0.18	0.20	−1.50	−0.72
0.75	0.18	0.14	0.34	0.31
1	0.15	0.05	0.29	0.19
1.5	0.03	−0.05	0.13	0.03
2	−0.11	−0.03	0.14	0.05
3	−0.26	−0.02	0.12	0.04
4	−0.11	0.06	0.16	0.08
6	−0.06	0.05	0.12	0.05
8	−0.03	0.05	0.11	0.05
12	0.00	0.05	0.10	0.04
16	−0.02	0.02	0.06	0.01
24	−0.01	0.01	0.05	0.00
32	−0.03	−0.01	0.02	−0.02
48	−0.04	−0.03	0.00	−0.04
64	−0.05	−0.05	−0.02	−0.05
Mean	−0.01	0.03	0.01	0.00
St. dev.	0.12	0.07	0.43	0.22
Maximum	0.18	0.20	0.34	0.31
Minimum	−0.26	−0.05	−1.50	−0.72

Table IV shows that Eq. (43) (D) is in good agreement with Sewell's (1970) formula. Sewell (1970) showed that his formula is in reasonable agreement with the sound insulation measurements of Utley (1986) in the Liverpool University facility on a limp lead panel, of surface density 17.6 kg/m^2, which measured 2.13×1.67 m^2. Thus Eq. (43) is in reasonable agreement with Utley's (1986) measurements.

Another experimental verification of the theory of this paper is the fact that the author and his students [Pavasovic (2006), Davy and Pavasovic (2006), Davy and Kannanaikkel John (2006), Fisher (2006), Davy (2007), Davy and Kannanaikkel John (2007), and Davy (2008)] have been able to use the theory of this paper with $n=1$, $w=1$, and $\beta=0$ to successfully predict the experimental directivity of the forced sound radiation from panels and openings.

VIII. CONCLUSIONS

The two dimensional strip model analytic approximation derived in this paper gives reasonable agreement with three dimensional numerical calculations. This agrees with Ljunggren (1991) whose two dimensional numerical calculations agree within ± 0.5 dB of the three dimensional calculations of Sato and Novak (1995). It also agrees with the experimental measurements of Roberts (1983), which show that the directivity of a rectangle depends strongly on its length in the direction of measurement but only weakly on its width at right angles to the direction of measurement. Thus this paper concludes that the radiation efficiency of a forced wave on a panel is mainly determined by the ratio of the panel's length, in the direction of propagation of the forced wave, to the wavelength of the sound in air and the angle of incidence of the forcing wave.

It should be pointed out that the forced radiation of sound from an opening can be treated as the forced radiation of sound from a panel because, for the same source size and angle of incidence of the forcing wave, the radiation depends only on the velocity distribution across the opening or panel. Consequently the results from this paper may be applied to predicting the radiation of sound from openings as well.

NOMENCLATURE

a = Half length of source
c = Speed of sound in air
f = Empirical limit of range of validity of Eqs. (10) and (11)
g = Cosine of angle of incidence
g_l = Theoretical limit of range of validity of Eqs. (10) and (11)
h = Inverse of radiation efficiency at grazing angles
H = Function
H_1 = Function
H_2 = Function
k = Wave number in air
k_b = Wave number in panel

L = Half shorter side length of rectangle
m = Constant
n = Empirical combining power
N = Number of sound sources
p = Sound pressure amplitude in air
q = Inverse of real part of low frequency radiation efficiency
r = Radius of sphere or hemisphere
S = Surface area
t = Time
U = Perimeter
w = Empirical correction factor for the limit of range of validity of equations (10) and (11)
W = Half longer side length of rectangle
x = Variable of integration, or aspect ratio of rectangle
y = Complement of angle of incidence
Z_{wf} = Fluid wave impedance of panel in air
α = Interpolation slope
β = Empirical offset correction for radiation efficiency at grazing incidence
δ = Half total phase change at observer
θ = Angle of radiation relative to normal
λ = Wavelength in air
ρ_0 = Ambient density of air
σ = Radiation efficiency
φ = Angle of incidence relative to normal
φ_l = Limiting angle of incidence relative to normal
ψ = Half change of phase across source
ω = Angular frequency
ω_c = Angular critical frequency

Cremer, L., and Heckl, M. (**1988**). *Structure-Borne Sound: Structural Vibrations and Sound Radiation at Audio Frequencies*, 2nd ed. translated and revised by E. E. Ungar (Springer-Verlag Company, New York).

Davy, J. L. (**2004**). "The radiation efficiency of finite size flat panels," *Acoustics 2004, Transportation Noise and Vibration—The New Millennium, Proceedings of the Annual Conference of the Australian Acoustical Society*, Gold Coast, Australia, 3–5 November, edited by M. J. Mee, R. J. Hooker, and I. D. M. Hillock, (Australian Acoustical Society, Castlemaine, Victoria, Australia), pp. 555–560.

Davy, J. L. (**2007**). "A model for predicting diffraction on a finite flat surface as a function of angle of incidence and surface size," *Global Approaches to Noise Control, Proceedings of the Inter-Noise 2007, the 36th International Congress and Exhibition on Noise Control Engineering*, Istanbul, Turkey, 28–31 August, (Turkish Acoustical Society, Istanbul, Turkey), Paper No. 21.

Davy, J. L. (**2008**). "The directivity of the forced radiation of sound from panels and openings including the shadow zone," in Proceedings of Acoustics'08, Paris, 29 June–4 July, Paper No. 11, pp. 3833–3838; J. Acoust. Soc. Am. **123**, 3499(A); Acta. Acust. Acust. **94**, S531(A).

Davy, J. L., and Kannanaikkel John, T. (**2006**). "The angular distribution of sound incident on a panel or opening (A)," J. Acoust. Soc. Am. **120**, 3187.

Davy, J. L., and Kannanaikkel John, T. (**2007**). "The directivity of the forced radiation of sound from panels and openings," in Proceedings of the 14 International Congress on Sound and Vibration (ICSV14), Cairns, Australia, 9–12 July, edited B. Randall, Paper No. 7.

Davy, J. L., and Pavasovic, V. (**2006**). "The directivity of sound radiated from a panel or opening excited by sound incident from the other side," in Proceedings of Acoustics 2006, Noise of Progress, First Australasian Acoustical Societies' Conference, Christchurch, New Zealand, 20–22 November, edited by T. McMinn, pp. 133–140.

Fisher, K. R. (**2006**). "Directivity of the sound transmission through finite sized apertures and panels," Bachelor of Applied Science (Applied Physics) with Honors, RMIT University, Melbourne, Australia.

Gradshteyn, I. S., and Ryzhik, I. M. (**1965**). *Table of Integrals, Series, and Products*, 4th ed., prepared by Yu. V. Geronimus and M. Yu. Tseytlin, translated from Russian by Scripta Technica Inc., edited by A. Jeffrey (Academic, New York).

Leppington, F. G., Broadbent, E. G., and Heron, K. H. (**1982**). "The acoustic radiation efficiency of rectangular panels," Proc. R. Soc. London, Ser. A **382**, 245–271.

Ljunggren, S. (**1991**). "Airborne sound insulation of thin walls," J. Acoust. Soc. Am. **89**, 2324–2337.

Lyon, R. H., and Maidanik, G. (**1962**). "Power flow between linearly coupled oscillators," J. Acoust. Soc. Am. **34**, 623–639.

Maidanik, G. (**1962**). "Response of ribbed panels to reverberant acoustic fields," J. Acoust. Soc. Am. **34**, 809–826; (**1975**). "Erratum: Response of ribbed panels to reverberant acoustic fields," J. Acoust. Soc. Am. **57**, 1552.

Novak, R. A. (**1992**). "Sound insulation of lightweight double walls," Appl. Acoust. **37**, 281–303.

Novak, R. A. (**1995**). "Radiation from partially excited plates," Acta Acust. **3**, 561–567.

Pavasovic, V. (**2006**). "The radiation of sound from surfaces at grazing angles of incidence," MS thesis, RMIT University, Melbourne, Australia.

Rindel, J. H. (**1975**). "Transmission of traffic noise through windows—Influence of incident angle on sound insulation in theory and experiment," Report No. 9, The Acoustics Laboratory, Technical University of Denmark, Lyngby, Denmark.

Rindel, J. H. (**1993a**), "Modelling the angle-dependent pressure reflection factor," Appl. Acoust. **38**, 223–234.

Rindel, J. H. (**1993b**). "Sound transmission through single layer walls," in Noise-93, Proceedings of the International Conference on Noise and Vibration Control, St. Petersburg, Russia, edited by M. J. Crocker, and N. I. Ivanov, pp. 63–68.

Roberts, J. (**1983**). "The prediction of directional sound fields," Transactions of the Institution of Engineers, Australia, Mechanical Engineering, **ME8**, 16–22.

Sewell, E. C. (**1970**). "Transmission of reverberant sound through a single-leaf partition surrounded by an infinite rigid baffle," J. Sound Vib. **12**, 21–32.

Thomasson, S.-I. (**1982**). "Theory and experiments on the sound absorption as function of the area," Report TRITA-TAK 8201, Royal Institute of Technology, Stockholm, Sweden.

Utley, W. A. (**1968**). "Single leaf transmission loss at low frequencies," J. Sound Vib. **8**, 256–261.

Development of a noise metric for assessment of exposure risk to complex noises

Xiangdong Zhu,[a] Jay H. Kim,[b] and Won Joon Song
Department of Mechanical Engineering, University of Cincinnati, Cincinnati, Ohio 45221-0072

William J. Murphy
National Institute for Occupational Safety and Health, Division of Applied Research and Technology, Engineering and Physical Hazards Branch, Hearing Loss Prevention Team, 4676 Columbia Parkway, MS C-27, Cincinnati, Ohio 45226-1998

Seongho Song
Department of Mathematical Sciences, University of Cincinnati, Cincinnati, Ohio 45221-0025

(Received 20 May 2008; revised 4 June 2009; accepted 5 June 2009)

Many noise guidelines currently use *A*-weighted equivalent sound pressure level L_{Aeq} as the noise metric and the equal energy hypothesis to assess the risk of occupational noises. Because of the time-averaging effect involved with the procedure, the current guidelines may significantly underestimate the risk associated with complex noises. This study develops and evaluates several new noise metrics for more accurate assessment of exposure risks to complex and impulsive noises. The analytic wavelet transform was used to obtain time-frequency characteristics of the noise. 6 basic, unique metric forms that reflect the time-frequency characteristics were developed, from which 14 noise metrics were derived. The noise metrics were evaluated utilizing existing animal test data that were obtained by exposing 23 groups of chinchillas to, respectively, different types of noise. Correlations of the metrics with the hearing losses observed in chinchillas were compared and the most promising noise metric was identified.
 [DOI: 10.1121/1.3159587]

PACS number(s): 43.60.Dh, 43.64.Tk, 43.66.Gf, 43.66.Vt [KA] Pages: 703–712

I. INTRODUCTION

Noise-induced hearing loss (NIHL) is the most common job-related illness affecting more than 11 million workers in the United States alone (EPA, 1981; NIOSH, 1998). Noise guidelines currently in use such as ISO-1999 (1990) and ANSI S3.44-1996 (R2006) (ANSI, 2006) recommend the noise exposure based on the equal energy hypothesis (EEH); therefore assume that the total *A*-weighted energy in a daily exposure can be used to predict NIHL. For example, ISO-1999 (1990) and NIOSH (1998) guidelines recommend the exposure limit of 85-dBA equivalent sound pressure level (SPL) (L_{Aeq}) as the action level and the 3-dB exchange rate. That is, for an 8-h workday, a time-weighted average of 85-dBA or less is expected to produce a material hearing impairment in about 10% of workers after a 40 year career working in noise (Prince *et al.*, 1998). For every 3-dB increase in exposure level, the allowable exposure duration is halved. The EEH-based approach is generally considered appropriate for steady-state noises but not for impulsive noises (Ahroon *et al.*, 1993b).

Typical workplaces are often subjected to a complex noise environment in which impulsive noises are embedded within a steady-state background noise. Current noise guidelines recommend integrating impulsive and continuous noises according to the EEH (ISO-1999, 1990; NATO, 1987; NIOSH, 1998). A number of animal exposure studies showed that the interaction effect between impulsive and broadband noises may actually exacerbate the NIHL (Blakeslee *et al.*, 1977; Hamernik *et al.*, 1987, 1974). For example, an exposure to a complex noise was observed to produce a much greater permanent threshold shift (PTS) and more extensive hair cell losses than an exposure to only an energy-equivalent continuous or impulsive noise alone would have caused (Blakeslee *et al.*, 1977; Hamernik *et al.*, 1974). Animal exposure studies strongly suggest the need for a more elaborate noise metric than L_{eq} for complex or impulsive noises (Ahroon *et al.*, 1993a, 1993b; Dunn *et al.*, 1991; Hamernik *et al.*, 1991; Hamernik and Qiu, 2001; Hamernik *et al.*, 2003a, 2003b; Hunt *et al.*, 1976; Martin, 1976; Nilsson *et al.*, 1983; Patterson, 1991; Roberto *et al.*, 1985; Starck and Pekkarinen, 1987; Starck *et al.*, 2003; Thiery and Meyer-Bisch, 1988; Voigt *et al.*, 1980).

Characteristics of the noise are typically described either in the time domain or frequency domain. However, *time* and *frequency* are interwoven concepts that change simultaneously (Kim *et al.*, 2007; Kulkarni *et al.*, 2004; Mallat, 1999; Zhu and Kim, 2006); therefore a fast transient event such as an impulsive or complex noise might better be characterized in the joint time-frequency (T-F) domain. The wavelet transform, which uses variable T-F atoms to break down signals, is well suited for analysis of transient signals. The analytic wavelet transform (AWT) is ideal to characterize noise and vibration signals (Kim *et al.*, 2007; Zhu and Kim, 2006). The AWT is a hybrid of the wavelet transform

[a]Present address: Parker-Haniffin Corp., 6035 Parkland Blvd., Cleveland, OH 44124–4141.
[b]Author to whom correspondence should be addressed. Electronic mail: jay.kim@uc.edu

TABLE I. Description of the noises that were used as the initial set of data for the correlation study in this paper. The data are from the chinchilla noise exposure studies conducted by Hamernik *et al.* (2003a, 2003b) and Qiu *et al.* (2006).

Noise index	Noise ID	Band type	Center frequency (kHz)	Band width (Hz)	Total SPL (dBA)	Time kurtosis	Peak SPL (dB)	Impulse probability of occurring in 750 ms window	Background level (dB)	Background level (dBA)
1	G-244	Three bands	1,2,4	400	100	25	115–128	0.6	98.5	95.5
2	G-249	Broadband	N/A	710–5680	100	33	115–129	0.6	95	91.5
3	G-250	Narrow band	N/A	1800–2200	100	21	114–128	0.6	98.6	95.5
4	G-251	Three bands	1,2,4	400	100	75	118–130	0.6	94.2	91
5	G-252	Three bands	1,2,4	400	100	53	123–127	0.6	96	92.6
6	G-253	Three bands	1,2,4	400	100	61	117–130	0.2	97.7	94
7	G-254	Three bands	1,2,4	400	100	25	15–128	Impact/1.5 s	97.7	94
8	G-255	Very broadband	N/A	100–10 000	100	25	115–129	0.6	98	94.5
9	G-259	Two bands	0.5,2	400	100	30	115–129	0.6	97.8	94.4
10	G-260	Three bands	1,2,4	400	100	39	115–129	0.6	97.2	93.5
11	G-261	Gaussian	N/A	N/A	100	3	N/A	N/A	N/A	N/A
12	G-263	Three bands	1,2,4	400	100	85–110	116–128	0.6	Very low	N/A
13	G-264	Three bands	1,2,4	400	100	12	116–126	0.6	100	97
14	G-265	Very broadband	N/A	100–10 000	100	105	127–132	0.1	Very low	N/A
15	G-266	Very broadband	N/A	100–10 000	100	15	113–127	0.6	99	96
16	G-268	Very broadband	N/A	100–10 000	100	65	128–133	0.1	99	96
17	G-269	Narrow band	N/A	1800–2200	100	75	114–129	0.6	Very low	N/A
18	G-270	Burst broadband	N/A	710–5680	100	27	104–115	0.6	96.7	91.3

and the Fourier transform that enables advantages of the wavelet transform while retaining familiar definitions such as frequency, amplitude, and SPL. The AWT is utilized as the basic signal analysis tool to characterize noise signals in this work.

This study aims at developing a new noise metric for more accurate assessment of the risk of complex noises. A good noise metric should relate the noise with the resulting hearing loss with high correlations. To identify such a metric, six basic forms of noise metrics were designed that reflect T-F characteristics of the noise in uniquely different ways. Fourteen noise metrics are obtained by varying parameters of the basic six forms. Correlations of the metrics with the hearing losses observed in chinchillas in the existing data from several animal exposure studies (Hamernik *et al.*, 2003a, 2003b, 1987, 1994; Hamernik and Ahroon, 1999; Hamernik and Qiu, 2000; Hamernik *et al.*, 2002) are compared to identify the most promising noise metric.

II. METHODS

A. Chinchilla noise exposure test data

Hamernik long proposed that a time-averaged metric such as the equivalent SPL (L_{eq}) is not sufficient to quantify the exposure hazard to complex noises (Blakeslee *et al.*, 1977; Hamernik *et al.*, 1987, 1994; Hamernik and Qiu, 2001). They designed various complex noises and conducted a series of NIHL tests on chinchillas using these noises. The complex noises were generated by combining various forms of impulse noise with a Gaussian continuous noise (Hamernik and Qiu, 2001). The total SPL was controlled to be the same for each study to give a standard platform for comparison. More details on the design of the noises can be found in Hamernik and Qiu, 2001; Hamernik *et al.*, 2003a, 2003b; and Lei *et al.*, 1994.

In the experiment, each group of chinchillas was exposed to, respectively, different types of noises for 5 days, 24 h/day, and then allowed to recover for 30 days. Auditory evoked potential pre-exposure (PRE), auditory evoked potential after 30 days post exposure (TS30), and auditory evoked potential after exposure (TS0) were measured. From these, the permanent threshold shift (PTS=TS30-PRE) and compound threshold shift (CTS=TS0-PRE) were calculated. Percentage of outer hair cell loss (%OHC) and percentage of inner hair cell loss (%IHC) were also carefully measured. The measurements were made over six octave band lengths of the basilar membrane with center frequencies at 0.5, 1, 2, 4, 8, and 16 kHz, that is, over the entire length of the basilar membrane.

The noises used in this study consisted of a Gaussian noise and various complex noises composed with variant levels of Gaussian background noise and impulses or bursts at different peak values, occurrence frequencies, and occurrence rates. The exposures of the chinchillas consist of several studies conducted by Hamernick *et al.* (Hamernik and Qiu, 2001; Hamernik *et al.*, 2003a, 2003b; and Lei *et al.*, 1994) to noises of three different levels: 90-, 95-, and 100-dBA. Table I describes 18 noise types for the exposures at 100-dBA. Table III describes 5 noise types at levels of 90- and 95-dBA.

TABLE II. Noise correlations of 14 metrics calculated by using the initial exposure data set composed of 18 animal groups exposed to 100-dBA noises. Each correlation is the correlation between the values of the given metric calculated for 18 noises and the PTS values of the 18 animal groups at each frequency. Notice that all metrics have very poor noise correlations at 0.5 and 2 kHz.

	Frequency (kHz)						
Metric index	0.5	1	2	4	8	16	Overall
1	−0.0072	0.533	−0.005	0.486	0.347	0.238	0.314
2	0.0526	−0.102	0.040	−0.203	−0.174	−0.243	−0.0511
3	0.0217	0.5412	0.0215	0.3057	0.1153	−0.152	−0.0094
4	−0.0131	0.634	0.1235	0.3201	0.0627	−0.219	−0.075
5	−0.0065	0.660	0.170	0.313	0.0153	−0.246	−0.102
6	−0.0021	0.666	0.196	0.302	−0.025	−0.256	−0.117
7	0.0008	0.665	0.213	0.293	−0.057	−0.268	−0.126
8	0.0069	0.652	0.248	0.268	−0.147	−0.286	−0.144
9	0.001	0.561	0.103	0.555	0.388	0.284	0.355
10	0.0136	0.591	0.138	0.570	0.555	0.306	0.356
11	0.046	0.602	0.156	0.559	0.646	0.318	0.351
12	0.080	0.607	0.167	0.546	0.696	0.336	0.349
13	0.172	0.607	0.193	0.525	0.727	0.520	0.361
14	−0.154	0.562	−0.324	0.268	0.243	−0.076	0.084

B. Analytic wavelet transform characterization of noise

Basic definitions and properties are briefly explained here; more details of the AWT can be found in Zhu and Kim, 2005 and Zhu and Kim, 2006. The AWT of a signal $f(t)$ of scale s is defined as

$$W_s(t) = \int_{-\infty}^{\infty} f(u)\psi_s^*(t-u)du, \tag{1}$$

where $\psi_s(u) = 1/s\, g(u/s)e^{j\eta(u/s)}$ is the wavelet function with scale s, $*$ indicates the complex conjugate, $j=\sqrt{-1}$, and a Gaussian function is chosen for $g(t)$ in this study, which is

$$g(t) = \frac{1}{(\sigma^2\pi)^{1/4}} e^{-t^2/2\sigma^2}, \tag{2}$$

where σ and η are parameters to be chosen. In this work $\sigma = 1.05$ and $\eta = 7.252$ are used, which effectively makes the AWT a real-time 1/3 octave frequency filter. Conducting the AWT as defined in Eq. (1) produces the time history of the 1/3 octave frequency component of the signal of the center frequency $\omega = \eta/s$. Equation (1) can be rewritten as

$$W_s(t) = \int_{-\infty}^{\infty} \frac{1}{s} f(u) g\left(\frac{t-u}{s}\right) e^{-j\eta[(t-u)/s]} du. \tag{3}$$

In this way, the AWT works like a transient Fourier transform. For example, by using the pressure time history measured in pascal for $f(u)$, the instantaneous SPL can be defined with the AWT result as follows:

$$\mathrm{SPL}(t,\omega) = 10\log_{10}\left(\frac{1}{2} W_s(t) W_s(t)^*\right) / {P_{\mathrm{ref}}}^2 dB, \tag{4}$$

where $P_{\mathrm{ref}} = 20\ \mu$Pa.

The AWT was applied to one of the noises used in the exposure study, G263 (Table I), for demonstration. The noises were recorded with a 48 000 Hz sampling rate for 5 min. A randomly selected 3-s period of the time history of the instantaneous pressure of the noise G263 is shown in Fig. 1.

The T-F characteristics of the noises obtained by applying the AWT technique are shown in Fig. 2 for the 3-s time history shown in Fig. 1. The plot was obtained by overlaying 1/3 octave time series of the noise obtained for center frequencies with 1/12 octave interval.

Each implementation of the AWT obtains the time history of the 1/3 octave SPL of the center frequency at $\omega = \eta/s$ (Zhu and Kim, 2006). Figure 3 shows six 1/3 octave time histories obtained for G263 at the center frequencies 0.5. 1, 2, 4, 8, and 16 kHz, at which NIHLs of chinchillas were measured. This set of six 1/3 octave time histories is used as the basic T-F information to calculate the noise metrics.

III. DESIGN OF NOISE METRICS

Six basic forms of noise metrics were designed to reflect T-F characteristics of noises in uniquely different ways,

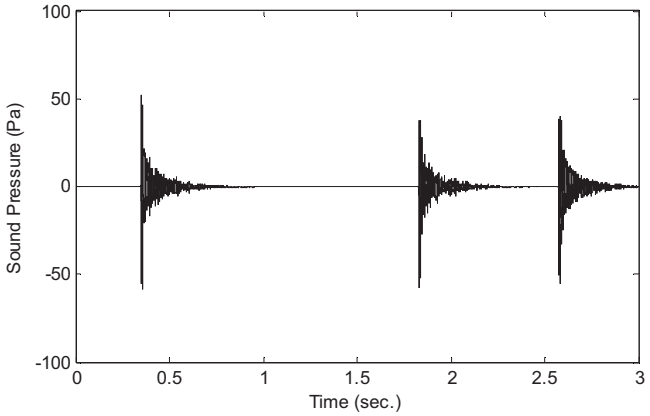


FIG. 1. Pressure time history of noise G263 for a 3-s period selected randomly out of 5 min period recorded.

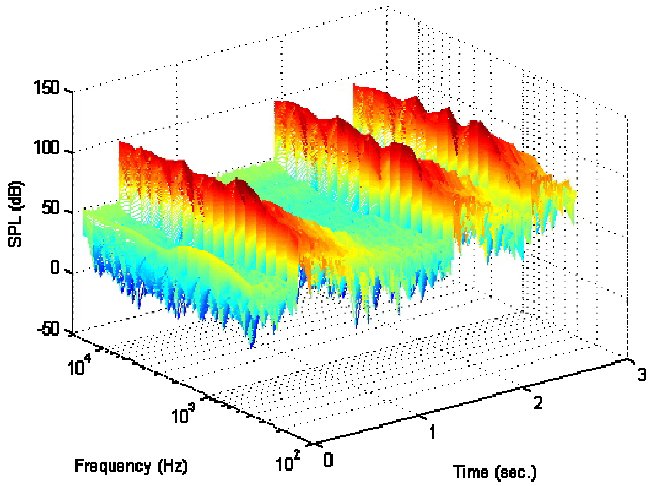


FIG. 2. (Color online) T-F representation of the pressure time history shown in Fig. 1 obtained by AWT. The height of the surface at a given frequency-time point indicates the SPL of the 1/3 octave frequency component centered at the frequency and at the time instant.

which will be varied to a total of 14 metrics. These metrics are calculated from the six 1/3 octave time histories; therefore are functions of frequency. The definitions of the metrics are as follows.

Metric 1: Equivalent SPL $L_{eq}(\omega)$. This metric has the same definition as the conventional L_{eq} except that it is calculated as a function of frequency. The time-averaged pressure $P_{eq}(\omega)$ is calculated as follows:

$$P_{eq}(\omega) = \sqrt{\frac{1}{T}\int_0^T (p(\omega,t))^2 dt}, \tag{5}$$

where $p(\omega,t)$ is the 1/3 octave pressure time histories obtained at the aforementioned six frequencies. $L_{eq}(\omega)$ is obtained as

$$L_{eq}(\omega) = 10\log_{10}\left(\frac{P_{eq}(\omega)^2}{P_{ref}^{\ 2}}\right). \tag{6}$$

Metric 2: The frequency domain kurtosis $\beta(\omega)$. The kurtosis, a statistical quantity that represents the impulsiveness of the event, is calculated from each 1/3 octave SPL time history. Kurtosis of a time series is defined as

$$\beta(\omega) = \frac{\mu_4}{\sigma^4}, \tag{7}$$

where, x is the time series, μ_4 is the fourth moment about the mean of x, and σ is the standard deviation of x.

The kurtosis obtained from the total sound pressure as a function of frequency was used as the metric in the noise exposure study by Hamernik and Qiu (2001), Hamernik *et al.* (2003a, 2003b), Lei *et al.* (1994), Lei and Hamernik (1995), and Qiu *et al.* (2006).

Metric 3: $L_{max}(\omega)$. Maximum SPL of each 1/3 octave time history is used as a metric based on the assumption that the hearing loss may depend on the maximum level of the frequency component. The 95% value of the SPL distribution histogram is taken as the maximum SPL.

Metrics 4–8: Dynamic SPL $L_d(\omega)$. The basic form of these metrics is defined as

$$L_d(\omega) = L_m(\omega) + K\Delta L(\omega), \tag{8}$$

where $L_m(\omega)$ is the mean value of the SPL of the noise, $\Delta L(\omega)$ is the dynamic fluctuation of the SPL defined as the difference between $L_{max}(\omega)$ (95% of the SPL distribution histogram) and $L_m(\omega)$, and K is a magnifying factor greater than 1. The design weights the dynamic component of the SPL more heavily than the static component based on the logic that the dynamic component is more detrimental than the static component to the failure of a dynamic system. Metrics

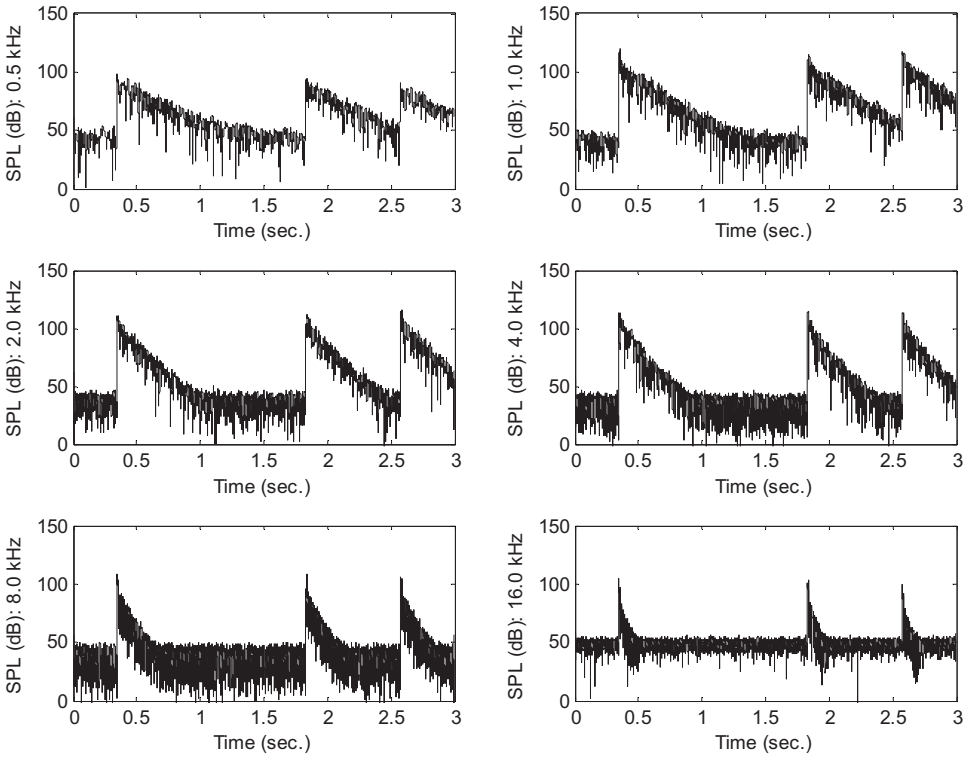


FIG. 3. Time histories of the 1/3 octave SPL components of noise G253. Each time history was obtained by applying the AWT to the noise with the center frequency at the frequency shown in the figure. For example, 0.5 kHz time history shown in the figure approximates the G263 noise that passed through a 1/3 octave filter of 0.5 kHz center frequency.

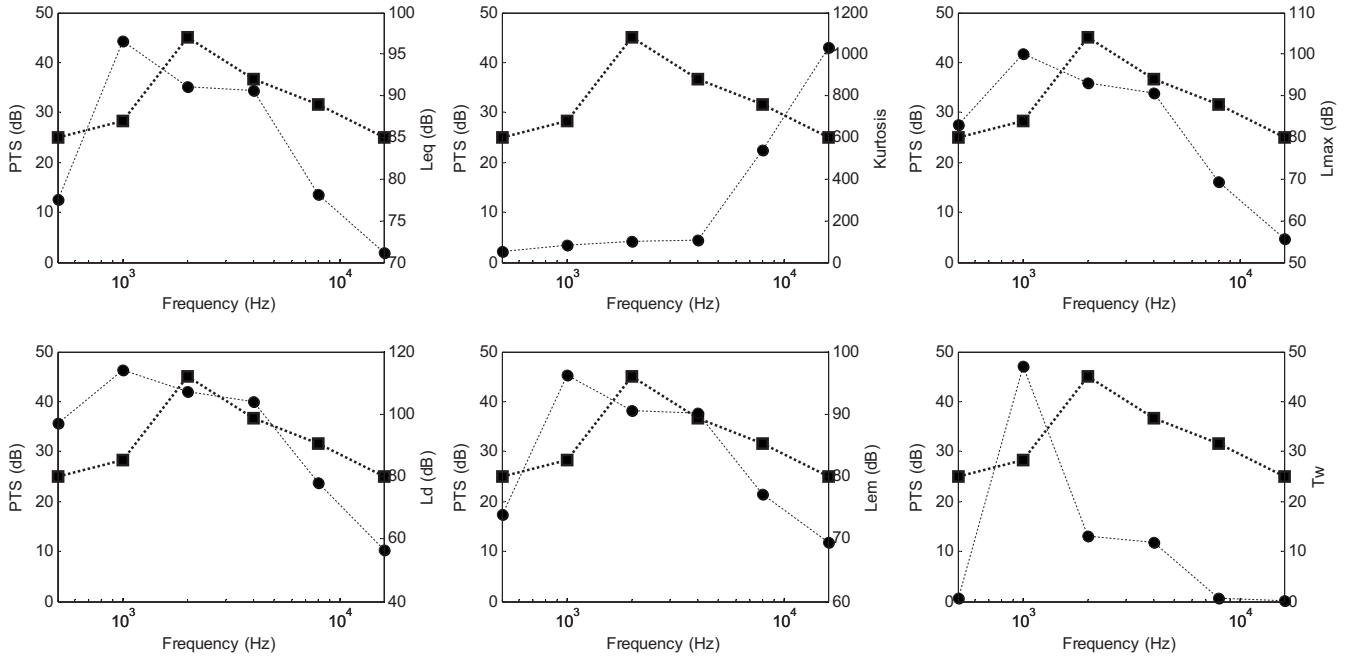


FIG. 4. Comparison of the median value of the measured PTS of the animals in group G263 and metric values calculated for the noise at six frequency points. Comparisons are shown for L_{eq}, kurtosis, and L_{max} (top, from left to right), and L_d ($K=2$), L_{em2}, and T_w (bottom, from left to right). Solid line indicates measured PTS and dashed line indicates metric values.

4–8 are defined by taking the values of $K=2$, 3, 4, 5, and 10.

Metrics 9–13: Modified equivalent SPL $L_{em}(\omega,\alpha)$. The basic definition is defined as follows:

$$P_{em}(\omega,\alpha)=\left[\frac{1}{T}\int_0^T \langle p(\omega,t)-p_o\rangle^{\alpha}dt\right]^{1/\alpha},$$

$$L_{em}(\omega,\alpha)\equiv L_{em\alpha}(\omega)=10\log_{10}\left(\frac{(P_{em}(\omega,\alpha))^{\alpha}}{P_{ref}^{\alpha}}\right), \tag{9}$$

where p_o is the threshold pressure, $p(\omega,t)$ is the pressure value of the 1/3 octave time history of SPL, and T is the averaging time. Higher value for the exponent α weights high pressure components more heavily. $\langle p(\omega,t)-p_0\rangle$ is a singular function defined as

$$\langle p(\omega,t)-p_0\rangle=0 \quad \text{if} \quad p(\omega,t)\le p_o,$$

$$\langle p(\omega,t)-p_0\rangle=p(\omega,t)-p_o \quad \text{if} \quad p(\omega,t)>p_o. \tag{10}$$

Equation (10) assumes that the noise contributes to the hearing loss only when its level exceeds the threshold pressure p_o. Currently, 0.282 Pa is used for p_o, which corresponds to the SPL of 80-dB, for all frequencies. Choice of p_o will have to be further studied. Metrics 9–13, respectively, are obtained by calculating with $\alpha=2$, 3, 4, 5, and 10.

Metric 14: Normalized weighted exposure time $\bar{T}_w(\omega)$. This definition is obtained by applying the 3-dB exchange rule to *each time interval* to each 1/3 octave time history for each unit time interval of size Δt, the sampling interval.

If $p(\omega,t)$ is the same as the threshold value p_0, the weighted time interval $\Delta t_{w,i}(\omega)=\Delta t$.

If $p(\omega,t)$ is higher than the threshold value by 3-dB, the weighted time interval $\Delta t_{w,i}(\omega)=2\Delta t$; by 6-dB, $\Delta t_{w,i}(\omega)=4\Delta t$; and so on.

If $p(\omega,t)$ in the interval is lower than the threshold value by 3-dB, the weighted time interval $\Delta t_{w,i}(\omega)=\Delta t/2$; by 6-dB, $\Delta t_{w,i}(\omega)=\Delta t/4$; and so on.

Finally, $\bar{T}_w(\omega)=\Sigma_{i=1}^{N}\Delta t_{w,i}(\omega)/T$, where T is the length of the time series and N is the number of time intervals.

IV. ANALYSIS OF PERMANENT THRESHOLD SHIFTS WITH THE PROPOSED NOISE METRICS

Four NIHL indicators, PTS, CTS, OHC, and IHC losses, were measured for each animal. In this study, PTS is selected as the NIHL parameter to study the correlation with the noise metrics because it is a direct indication of the hearing loss.

Six of the 14 metrics are plotted with the PTS data obtained from G263 group (see Table I) in Fig. 4 as an example. It is seen visually that $L_{max}(\omega)$, $L_d(\omega)$, and $L_{em2}(\omega)$ give good correlations with PTS for this particular group.

Figure 5 shows box plots of the PTS data measured at 6 frequencies from 18 chinchilla groups. Boxes represent interquartile ranges, horizontal lines represent the median, whiskers represent the largest and smallest values, and "+" symbols represent outliers defined as the points outside of 1.5 box lengths from the end of the boxes. The measured PTS data had very large statistical variations. In some cases, the animals in the same group showed more than 30-dB difference in the measured PTS value. It is known that typical individual variability in animals can be very high for unknown reasons (Cody and Robertson, 1983). Because of the high statistical variations, median values are used instead of mean values in the ensuing correlation study.

V. CORRELATION STUDY OF METRICS

The correlation between two sets of data $\{x\}$ and $\{y\}$ is defined as follows:

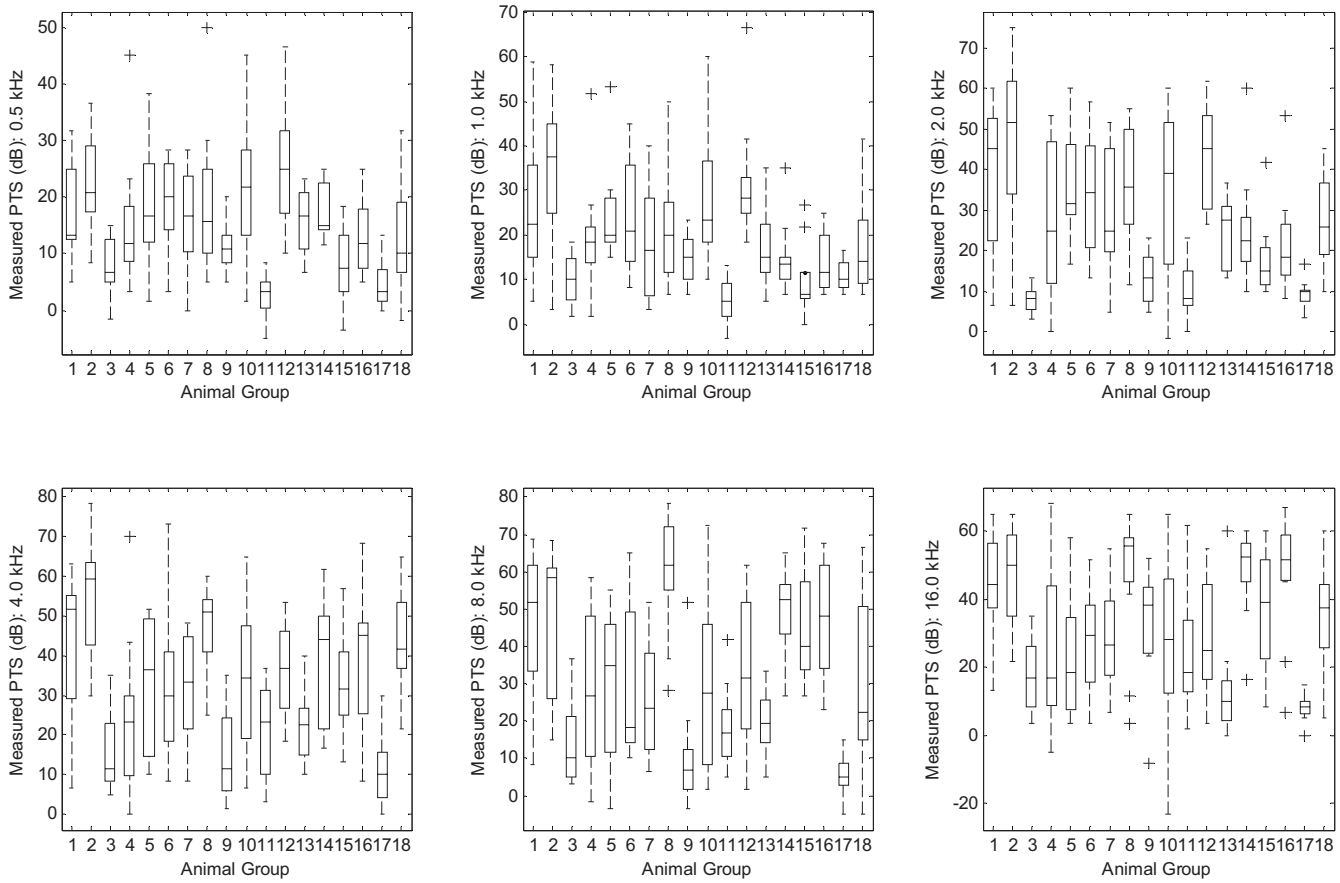


FIG. 5. Box plots of measured PTS of 18 groups of chinchillas at 0.5, 1, 2, 4, 8, and 16 kHz. Boxes represent inter-quartile ranges, horizontal lines represent the median, whiskers represent the largest and smallest values, and + symbols represent outliers defined as the points outside of 1.5 box lengths from the end of the boxes.

$$\rho_{xy} = \frac{E[(x-m_x)(y-m_y)]}{\sigma_x \sigma_y}, \tag{11}$$

where $\sigma_x{}^2 = E[x^2] - (E[x])^2$ and $\sigma_y{}^2 = E[y^2] - (E[y])^2$ are variances of $\{x\}$ and $\{y\}$, E stands for expectation, m_x is the mean value of x, and m_y is the mean value of y. $|\rho_{xy}| = 1$ indicates a perfect correlation and $\rho_{xy} = 0$ indicates no correlation.

Initially, correlation study was conducted using the exposure data obtained by exposing chinchillas to 18 types of noise shown in Table I. As it was mentioned, the median value of the PTS of animals in each noise group was used for the study. The metric values and the PTS data are obtained for 18 noises (or animal groups) at 6 frequency points, therefore, in 18×6 matrices. Let matrix $m_i(g_1:g_{18}, f_1:f_6)$ be the ith noise metric calculated for 18 noises at 6 frequencies and matrix PTS$(g_1:g_{18}, f_1:f_6)$ the PTS of 18 animal groups at 6 frequencies. Three correlations between the metric and measured PTS are defined in this study as follows.

Frequency correlation indicates how well the metric and PTS are correlated as functions of frequency. For example, in Fig. 4, which compares PTS and six metrics as functions of frequency, L_{eq} and L_{em2} have good frequency correlations. Frequency correlations are calculated for each of the 14 metrics for 18 metrics from 6 pairs of the metric-PTS data. Therefore, 14×8 frequency correlations are obtained, which are shown in Fig. 6 in a graphical form.

Noise correlation indicates how well the metric and PTS of animal groups are correlated for the 18 noises at each frequency. 6 noise correlations for each of the 14 metrics, therefore, 6×14 noise correlations are calculated. For example, the noise correlation of L_{em2} (ninth metric) at 1 kHz (second frequency) is calculated from 18 pairs of data

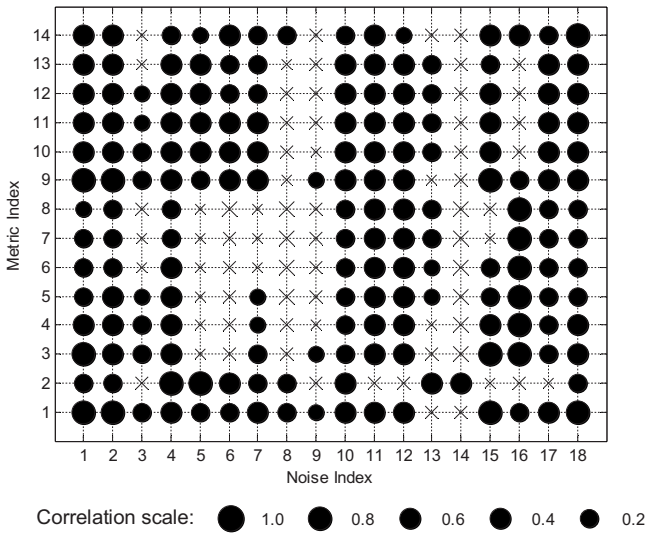


FIG. 6. Frequency correlation of noise metrics and measured PTS as functions of frequency. The size of a dot represents the correlation values as the scale shows. For example, the dot corresponding to noise index 5 (G252) and metric index 9 (L_{em2}) represents the frequency correlation of L_{em2} in animal group G252. The two metrics that showed the highest average of the frequency correlations are metric 1 (L_{eq}) and metric 9 (L_{em2}). x indicates a negative correlation.

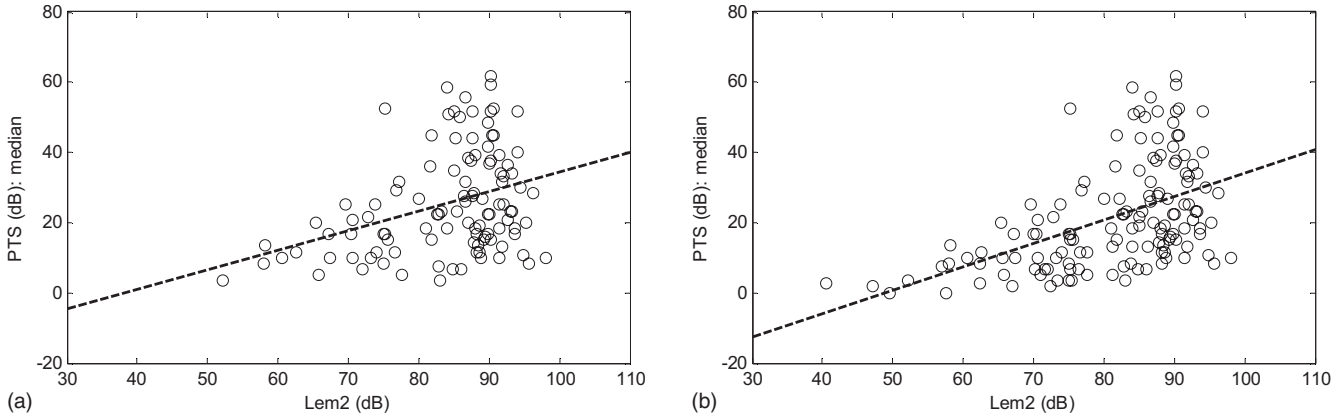


FIG. 7. PTS-L_{em2} scatter plots showing the overall correlation of L_{em2}: (a) plot of the initial data (108 points:18 groups, 6 frequency points) and (b) plot of the expanded data set (138 points: 23 groups, 6 frequency points).

$[m_9(g_1:g_{18},f_2)$; PTS$(g_1:g_{18},f_2)]$ by using Eq. (11).

Overall correlation indicates the overall performance of the given metric. The overall correlation is calculated from the combined data set of the above two cases. The correlation is calculated from Eq. (11) using all $108(18\times6)$ pairs of data $[m_9(g_1:g_{18},f_1:f_6)$; PTS$(g_1:g_{18},f_1:f_6)]$. A single correlation value is obtained for each metric.

Table II shows noise correlations and overall correlations calculated from the initial exposure data obtained from 18 animal groups shown in Table I.

A. Frequency correlations

Considering the tonotopic response of the basilar membrane, good correlations between the frequency dependencies of the NIHL and the noise metric can be expected. Figure 6 shows frequency correlations of the 14 metrics calculated for 18 noises in a dot map, in which the size of a dot represents the correlation values. A larger dot represents a better correlation and "x" represents a negative correlation. Noise index in the horizontal axis indicates the noise to which the animals were exposed (see Table I), and metric index in the y-axis indicates the aforementioned 14 noise metrics. For example, the dot corresponding to noise index 3 and metric index 9 represents the correlation of noise metric 9 $[L_{em}(\omega,\alpha=2)$ with noise 3 (G250)]. An ideal metric would be highly correlated with all 18 noises. For some unknown reason, no metric gave good correlations for noises 8, 9, and 14 (G255, G259, and G265). Examination of the measured NIHL data did not reveal any abnormality for these groups; however, the examination was of limited nature because the data are from a past study.

The two metrics that show the highest frequency correlation values are metric 1 $(L_{\mathrm{eq}}(\omega))$ and metric 9 $(L_{em}(\omega,\alpha=2))$. All five modified equivalent SPLs $[L_{em}(\omega)$ with $\alpha=2,3,4,5,10]$ exhibited correlation coefficients close to the top two metrics. The frequency correlation can have a large statistical error because it is calculated from only six data points. Also, it does not represent how well the metric represents risks of different noises. Therefore the frequency metric is not used as the indicator of the quality of the metric.

B. Noise correlations and overall correlations

The noise correlation and overall correlation defined previously indicate the ability of the noise metric to represent NIHL risks of different types of noise. Therefore the noise correlation is used to compare the metrics. Table II shows noise correlations and overall correlations calculated for the 14 metrics using the initial exposure data given in Table I. Metric 9 $(L_{em}(\omega),\alpha=2)$ and metric 11 $(L_{em}(\omega),\alpha=4)$ give the best overall correlation value. It is seen that all metrics show very poor noise correlations at 0.5 and 2 kHz, implying none of the metrics is able to assess the risk of noises at these frequencies. This led to a further investigation of the data.

When the values of 14 metrics calculated at 6 frequencies from the 18 noises were compared, it was found that the range of the metric values at 0.5 and 2 kHz was significantly smaller than those at other frequencies. This may have made the data ill-conditioned in calculation of the noise correlation, inducing large errors in estimated correlation values. Taking an extreme example, if all chinchillas were exposed to the noises that have exactly the same metric value, the PTS—metric scatter plot that uses metric as the abscissa (see Fig. 7 for an example)—will form a horizontal line, thus a zero-valued correlation will be resulted.

Based on the above hypothesis, the chinchilla exposure data obtained using lower level noises were added to expand the range of the data. Table III shows the information of the additional test noises. The additional data were obtained from 61 chinchillas divided into 5 groups exposed to various types of noise of 90- and 95-dBA. Using the expanded data obtained from a total of 275 chinchillas exposed to 23 different noises is expected to improve accuracy of estimated correlations. Figure 7(a) shows PTS-L_{em2} scatter plots of L_{em2} obtained from the initial data (108 points:18 groups, 6 frequency points) and in Fig. 7(b) from the expanded data set (138 points: 23 groups, 6 frequency points). The overall correlation coefficients were calculated to be 0.355 [Fig. 7(a)] and 0.498 [Fig. 7(b)].

The noise and overall correlations calculated from this expanded data are shown in Table IV. The correlations are generally higher, especially at 0.5 and 2 kHz, compared to those reported in Table II. It is seen that the traditional metric L_{eq} shows quite good noise correlations. All L_{em} metrics

TABLE III. Description of the noises that were used as the additional data for the correlation study in this paper. The data are from a new, unpublished chinchilla noise exposure study conducted by Hamernik and Qiu (2001).

Noise index	Noise ID	SPL [db(A)]	Band type	Band width (Hz)	Time kurtosis	Peak SPL (dB)	Probability of impulse in 750 ms	No. of animals in the group
19	G-247	90	Gaussian	N/A	3	N/A	N/A	12
20	G-248	90	Three bands	400	32	104–118	0.6	11
21	G-256	90	Very broadband	100–10000	35	105–120	0.6	11
22	G-257	95	Gaussian	N/A	3	N/A	N/A	15
23	G-258	95	Three bands	400	41	109–124	0.6	12

show good noise correlations at all frequencies, and a good overall correlation. Correlations of L_{em} metrics are expected to be further improved by choosing more proper values of the threshold level p_o [see Eq. (9)], for which an arbitrary value of 80-dB was taken in this study. Among the five L_{em} variations (metrics 9–13), L_{em2} has an important advantage. L_{em2} can be interpreted the same way as the traditional metric L_{eq} for high level noises because of the way it is defined. For example, L_{em2} will change by 3-dB when L_{eq} of the noise changes by 3-dB if the noise has higher SPL than the threshold level. Therefore, L_{em2} is identified as the most promising metric to be used to develop an advanced noise risk assessment procedure.

VI. DISCUSSION AND CONCLUSION

Fourteen noise metrics derived from six basic designs were studied to identify the most promising noise metric that can be used for more accurate assessment of risks of complex noises. The metrics were designed so that T-F characteristics of the noise are reflected in uniquely different ways. T-F characteristics of the noises are obtained as a set of time histories of 1/3 octave SPL of the noise at 6 frequency points by applying the analytic wavelet transform. Initially, the 14 noise metrics were evaluated by their correlations with existing animal noise exposure study data obtained by exposing 18 groups of chinchillas to, respectively, different types of noise of 100-dBA overall SPL. The correlation study was conducted by statistically comparing the metric values calculated from the noise and the PTS measured in the chinchillas exposed to the noise.

Three types of correlations of the noise metrics were used, which were the frequency correlation (correlation as functions of frequencies), the noise correlation (correlation across noise groups), and the overall correlation. All versions of modified equivalent SPL, $L_{em}(\omega,\alpha)$, showed good frequency correlations, which have similar definitions to $L_{\text{eq}}(\omega)$ but counting the noise only its instantaneous level is higher than a threshold value.

In the correlation study with the initial data set, it was found that all noise metrics showed very poor noise correlations at 0.5 and 2 kHz. Examining the metric values calculated from the initial data set composed of 18 noises; it was seen that metric values at these two frequencies were in relatively small ranges, which may have made calculation of the noise correlations ill-conditioned. The data set was expanded by adding the exposure data obtained with lower level noises to increase the metric data range. The exposure data from 5 groups of 61 chinchillas exposed to noises of 90- and 95-dBA were added to the initial data. Noise correlations calculated from these expanded data showed significantly higher

TABLE IV. Noise correlations of 14 metrics calculated by using the expanded exposure data. The expanded data were obtained by adding the data from three animal groups exposed to 90-dBA noises and two animal groups exposed to 90-dBA noises. Correlations have become higher in general, especially at 0.5 and 4 kHz.

	Frequency (kHz)						
Metric index	0.5	1	2	4	8	16	Overall
1	0.207	0.658	0.430	0.632	0.449	0.327	0.453
2	0.236	0.141	0.274	0.0169	−0.066	−0.084	0.046
3	0.182	0.615	0.311	0.384	0.146	−0.157	0.052
4	0.191	0.569	0.337	0.360	0.081	−0.226	−0.032
5	0.183	0.658	0.344	0.327	0.024	−0.253	−0.068
6	0.172	0.647	0.346	0.300	−0.025	−0.268	−0.087
7	0.163	0.637	0.347	0.280	−0.063	−0.277	−0.099
8	0.139	0.604	0.345	0.231	−0.167	−0.295	−0.124
9	0.287	0.674	0.491	0.662	0.511	0.457	0.498
10	0.300	0.692	0.499	0.656	0.614	0.532	0.502
11	0.328	0.697	0.502	0.645	0.669	0.574	0.499
12	0.356	0.700	0.504	0.636	0.698	0.608	0.498
13	0.427	0.697	0.509	0.623	0.696	0.717	0.505
14	0.0212	0.649	−0.055	0.470	0.378	0.220	0.241

values, especially at 0.5 and 2 kHz. All variations of L_{em} had good overall and noise correlations; therefore may serve as good noise metrics. Among these metrics, L_{em2} is considered the most promising noise metric because it will be easier to use as its definition is compatible with the definition of L_{eq}.

Practicality of the approach may be questioned in relation to developing noise guidelines. For example, adoption of the new noise metric requires additional processing, that is, digital recording of the noise and its computer analysis. The problem is believed to be a minor issue considering that the modern computing becomes ever more powerful and affordable. The applicability of the noise metric developed for chinchillas to human noise guidelines is a more complex issue; although, chinchillas and human have very similar audiometric characteristics. For example, the chinchilla ear may have different susceptibilities than the threshold sensitivity suggests. This issue will have to be further addressed. Development of a strategy to utilize the new noise metric to develop noise guidelines is being studied by the authors. The authors are also developing a method to reflect response characteristics of the ear in the design of noise metric. Specifically, the simulated response of the stapes or basilar membrane is being used to calculate noise metrics. In addition to improve performance of the noise metric, the approach may provide some insights into the process of hearing loss.

ACKNOWLEDGMENTS

This project was supported by the National Institute for Occupational Safety and Health, Grant No. R21 OH008510. The authors thank Roger Hamernik and Wei Qiu at the State University of New York at Plattsburgh for providing chinchilla noise exposure study data and advice in interpreting the data.

Ahroon, W. A., Davis, R. I., and Hamernik, R. P. (**1993a**). "The role of tuning curve variables and threshold measures in the estimation of sensory cell loss," Audiology **32**, 244–259.

Ahroon, W. A., Hamernik, R. P., and Davis, R. I. (**1993b**). "Complex noise exposures: An energy analysis," J. Acoust. Soc. Am. **93**, 997–1006.

ANSI (**2006**). "Determination of occupational noise exposure and estimation of noise-induced hearing impairment," S3.44-1996 (R2006),American National Standards Institute.

Blakeslee, E. A., Hynson, K., Hamernik, R. P., and Henderson, D. (**1977**). "Interaction of spectrally mismatched continuous and impulse-noise exposures in the chinchilla," J. Acoust. Soc. Am. **61**, S59.

Cody, A. R., and Robertson, D. (**1983**). "Variability of noise-induced damage in the guinea pig cochlea: Electrophysiological and morphological correlates after strictly controlled exposures," Hear. Res. **9**, 55–70.

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.

EPA (**1981**). "Noise in America: The extent of the noise problem," USEPA (ONAC) Report No. 550/9-81-101, U.S. Environmental Protection Agency, Office of Noise Abatement and Control, Washington, DC.

Hamernik, R. P., and Ahroon, W. A. (**1999**). "Susceptibility of the noise-toughened auditory system to noise-induced trauma," Hear. Res. **132**, 140–148.

Hamernik, R. P., Ahroon, W. A., Davis, R. I., and Lei, S. F. (**1994**). "Hearing threshold shifts from repeated six-hour daily exposure to impact noise," J. Acoust. Soc. Am. **95**, 444–453.

Hamernik, R. P., Ahroon, W. A., Henderson, D., and Salvi, R. J. (**1987**). "The interaction between continuous and impulse noise: Frequency effects," The Second International Conference on the Combined Effects of Environmental Factors, Koyei, Kanazawa, Japan.

Hamernik, R. P., Ahroon, W. A., and Hsueh, K. D. (**1991**). "The energy spectrum of an impulse: Its relation to hearing loss," J. Acoust. Soc. Am. **90**, 197–204.

Hamernik, R. P., Ahroon, W. A., Patterson, J. H., Jr., and Qiu, W. (**2002**). "Relations among early postexposure noise-induced threshold shifts and permanent threshold shifts in the chinchilla," J. Acoust. Soc. Am. **111**, 320–326.

Hamernik, R. P., Henderson, D., Crossley, J. J., and Salvi, R. J. (**1974**). "Interaction of continuous and impulse noise: Audiometric and histological effects," J. Acoust. Soc. Am. **55**, 117–121.

Hamernik, R. P., and Qiu, W. (**2000**). "Correlations among evoked potential thresholds, distortion product otoacoustic emissions and hair cell loss following various noise exposures in the chinchilla," Hear. Res. **150**, 245–257.

Hamernik, R. P., and Qiu, W. (**2001**). "Energy-independent factors influencing noise-induced hearing loss in the chinchilla model," J. Acoust. Soc. Am. **110**, 3163–3168.

Hamernik, R. P., Qiu, W., and Davis, B. (**2003a**). "Cochlear toughening, protection, and potentiation of noise-induced trauma by non-Gaussian noise," J. Acoust. Soc. Am. **113**, 969–976.

Hamernik, R. P., Qiu, W., and Davis, B. (**2003b**). "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.

Hunt, W. J., Hamernik, R. P., and Henderson, D. (**1976**). "The effects of impulse level on the interaction between impulse and continuous noise," Trans.-Am. Acad. Opthalmol. Otolayngol., Sec. Ophthalmol. **82**, 305–308.

ISO-1999 (**1990**). *Acoustics—Determination of Occupational Noise Exposure and Estimation of Noise-Induced Hearing Impairment*, 2nd ed. (International Standards Organization, Geneva, Switzerland).

Kim, J., Welcome, D. E., Dong, R. G., Joon Song, W., and Hayden, C. (**2007**). "Time–frequency characterization of hand-transmitted, impulsive vibrations using analytic wavelet transform," J. Sound Vib. **308**, 98–111.

Kulkarni, P., Zhu, X., Kim, J., and Hayden, C. (**2004**). "Time-frequency analysis of high-intensity, fast transient noise from power tools by analytic wavelet transform," NoiseCon 2004, Baltimore, MD.

Lei, S. F., Ahroon, W. A., and Hamernik, R. P. (**1994**). "The application of frequency and time domain kurtosis to the assessment of hazardous noise exposures," J. Acoust. Soc. Am. **96**, 1435–1444.

Lei, S. F., and Hamernik, R. P. (**1995**). "Wavelet transform and frequency domain kurtosis: Application to assessment of hearing hazard from noise exposure," 1995 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, 15–18 October, pp. 19–21.

Mallat, S. (**1999**). *A Wavelet Tour of Signal Processing*, 2nd ed. (Academic, New York).

Martin, A. M. (**1976**). "The equal energy concept applied to impulse noise," *The Effects of Noise on Man*, edited by D. Henderson, R. P. Hamernik, D. S. Dosanjh, and J. H. Mills (Raven, New York).

NATO (**1987**). "The effects of impulse noise," Document AC/243/(PANEL8/RSG.6)D/9, Research Study Group RSG6/PANEL8, North Atlantic Treaty Organisation, Brussels.

Nilsson, R., Dunn, D. E., Erlandsson, B., Brenner, J., Hakanson, H., and Ivarson, A. (**1983**). "Conclusions from animal experiments on the effects of steady-state and impulse noise," Noise as a Public Health Problem: Proceedings of the Fourth International Congress, Milano, Italy.

NIOSH (**1998**). "Criteria for a recommended standard—Occupational noise exposure," DHHS (NIOSH) Publication No. 98-126, U.S. Department of Health and Human Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, Cincinnati, OH.

Patterson, J. H., Jr. (**1991**). "Effects of peak pressure and energy of impulses," J. Acoust. Soc. Am. **90**, 205–208.

Prince, M. M., Stayner, L. T., Smith, R. J., and Gilbert, S. J. (**1998**). "Response to "Comments on ′A re-examination of risk estimates from the NIOSH Occupational Noise and Hearing Survey′ " [J. Acoust. Soc. Am. **103**, 2734 (**1998**)]," J. Acoust. Soc. Am. **103**, 2736–2739.

Qiu, W., Hamernik, R. P., and Davis, B. (**2006**). "The kurtosis metric as an adjunct to energy in the prediction of trauma from continuous, non-Gaussian noise exposure," J. Acoust. Soc. Am. **120**, 3901–3906.

Roberto, M., Hamernik, R. P., Salvi, R. J., Henderson, D., and Milone, R. (**1985**). "Impact noise and the equal energy hypothesis," J. Acoust. Soc. Am. **77**, 1514–1520.

Starck, J., and Pekkarinen, J. (**1987**). "Industrial impulse noise: Crest factor as an additional parameter in exposure evaluation," Appl. Acoust. **20**, 263–274.

Starck, J., Toppila, E., and Pyykko, I. (**2003**). "Impulse noise and risk criteria," Noise Health **5**, 63–73.

Thiery, L., and Meyer-Bisch, C. (**1988**). "Hearing loss due to partially impulsive industrial noise exposure at level between 87 and 90 dB(A)," J. Acoust. Soc. Am. **84**, 651–659.
Voigt, P., Godenhielm, B., and Ostlund, E. (**1980**). "Impulse noise-measurement and assessment of the risk of noise induced hearing loss," Scand. Audiol. Suppl. **12**, 319–325.
Zhu, X., Kim, J. H. (**2005**). "Application of the Analytic Wavelet Transformation for Time-Frequency Analysis of Impulsive Sound Signals," in *SAE 2005 Noise and Vibration Conference* (SAE International, Grand Traverse, MI, USA).
Zhu, X., and Kim, J. (**2006**). "Application of analytic wavelet transform to analysis of highly impulsive noises," J. Sound Vib. **294**, 841–855.

Travel-time sensitivity kernels versus diffraction patterns obtained through double beam-forming in shallow water

Ion Iturbe[a)]
GIPSA Laboratory, INPG–CNRS, 961 rue de la Houille Blanche, Domaine Universitaire, BP 46, 38402 Saint Martin d'Heres, France

Philippe Roux and Jean Virieux
LGIT, Universite Joseph Fourier–CNRS, 1381 rue de la Piscine, Saint Martin d'Heres, 38400 France

Barbara Nicolas
GIPSA Laboratory, INPG–CNRS, 961 rue de la Houille Blanche, Domaine Universitaire, BP 46, 38402 Saint Martin d'Heres, France

(Received 5 August 2008; revised 22 May 2009; accepted 27 May 2009)

In recent years, the use of sensitivity kernels for tomographic purposes has been frequently discussed in the literature. Sensitivity kernels of different observables (e.g., amplitude, travel-time, and polarization for seismic waves) have been proposed, and relationships between adjoint formulation, time-reversal theory, and sensitivity kernels have been developed. In the present study, travel-time sensitivity kernels (TSKs) are derived for two source-receiver arrays in an acoustic waveguide. More precisely, the TSKs are combined with a double time-delay beam-forming algorithm performed on two source-receiver arrays to isolate and identify each eigenray of the multipath propagation between a source-receiver pair in the acoustic waveguide. A relationship is then obtained between TSKs and diffraction theory. It appears that the spatial shapes of TSKs are equivalent to the gradients of the combined direction patterns of the source and receiver arrays. In the finite-frequency regimes, the combination of TSKs and double beam-forming both simplifies the calculation of TSK and increases the domain of validity for ray theory in shallow-water ocean acoustic tomography. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3158922]

PACS number(s): 43.60.Fg, 43.60.Rw [WLS] Pages: 713–720

I. INTRODUCTION

The resolution limit of travel-time tomography has been studied from various aspects.[1–3] This investigation essentially relies on the specific, maybe paradoxical, nature of travel-times, as extracted from time-series recordings. Once picked, travel-times lose the frequency information of the time series. For example, in seismology, choosing times from high-frequency impulsive seismograms or from broad-band low-frequency seismograms will certainly have an impact on the tomographic resolution. However, the frequency information is not used in the travel-time tomography machinery based on ray theory (for example, see Ref. 4). *Ad hoc* procedures for introducing frequency information have been designed[5] with the so-called fat-ray concept, based on reconstruction assumptions. The more physical concept of the wave path, as related to the wave propagation properties, was introduced by Woodward,[6] which is closely related to Fresnel tomography in optics.[7] In recent years, this finite-frequency influence has been systematically investigated for the different observables (i.e., time, polarization, amplitude, and anisotropy) in different studies, suggesting that higher resolution images can be obtained from this improved description of wave propagation physics (see Ref. 8 for a general review).

Based on single-scattering effects, sensitivity kernels have been introduced and different computational techniques have been devised from ray theory as paraxial theory[9] or exact ray theory[10] to numerical tools.[11] Different studies[12] have questioned the differential techniques used for the construction of these sensitivity kernels, with the emphasis on the so-called banana-doughnut paradox: for travel-times, the sensitivity kernel is zero on the ray connecting the source and the receiver. Other studies[13] have shown that the travel-time tomographic problem with the specific density of stations and sources encountered in seismology prevents an improvement of resolution. Overcoming these limitations of data quality requires a denser deployment of sources and receivers, which can be expensive.

Similar to seismic studies, consideration of arrays of sources and receivers is a classical approach in underwater acoustics.[14] Wave-propagation problems can lead to similar features as in geophysics, and we would like to investigate the effects of the finite size of the source and/or receiver arrays in underwater tomographic reconstructions. The concept of sensitivity kernels has been applied recently to this field,[15] and quite exciting theoretical and experimental investigations have led to fruitful achievements with links to time-reversal theory,[16,17] adjoint methods,[11] and acoustic and seismic imagings[18] or medical imaging.[19–21]

In the present study, the relationships between travel-time sensitivity kernel (TSK) reconstruction and diffraction theory in the context of shallow underwater acoustics is de-

[a)]Author to whom correspondence should be addressed. Electronic mail: ion.iturbe@gipsa-lab.inpg.fr

fined when the transfer function of the waveguide is recorded between source and receiver arrays. We show that the spatial shapes of TSKs are equivalent to the gradient of the combined diffraction pattern of the arrays. This relationship is exact when working with a point-to-point approach (using only one source and one receiver), and it becomes more complex and approximate when TSKs are calculated between two source and receiver arrays.

In a shallow-water environment, array processing using source and/or receiver arrays is necessary to improve the separation of the different ray paths. One standard array processing method is time-delay beam-forming on the receiver array, to separate the ray paths according to their receiver angles.[22] Recently, Roux *et al.*[23] proposed a more sophisticated time-delay double beam-forming (DBF) algorithm, based on spatial reciprocity, which takes advantage of both receiver and source arrays. The DBF algorithm can be applied when the entire transfer matrix is measured between each pair of source-receiver transducers. DBF consists of transforming the three-dimensional (3D) data space from source depth, receiver depth, and time into a new 3D space that is related to ray propagation, described by the beam-formed variables: source angle, receiver angle and time. As a consequence, every eigenray of the multipath propagation for a source-receiver pair is identified and isolated through DBF according to the receiver and source angles.

In their very recent study, Roux *et al.*[23] went one step further. Every eigenray isolated through DBF processing became free from any interference effects due to multipath propagation. Furthermore, DBF processing provides array gain and robustness, since every DBF eigenray arises from the summation of a large number of time-delayed source-to-receiver signals. Thus, both the amplitude and phase of the DBF eigenray can be followed as a function of dynamic ocean fluctuations, when, for example, internal waves locally perturb the sound-speed profile (see Fig. 7 of Ref. 23).

The stability and robustness of DBF processing can lead to important consequences for future studies relating to ocean acoustic tomography. To date, travel-time tomography has mainly been performed from echo arrival peaking through point-to-point measurements. Indeed, only the envelope of the demodulated signal was a robust observable, regarding signal-to-noise ratio issues and rapidly changing ocean fluctuations (e.g., gravity waves at the ocean surface). The use of source-receiver arrays now allows the travel-time fluctuations to be measured as phase changes in the DBF signal, providing travel-time measurements with greater accuracy. Indeed, travel-time change measured through the phase has an accuracy driven by the carrier frequency F_c of the signal, while travel-time changes measured through the envelope of the demodulated signal depend on the frequency bandwidth $\delta f < F_c$.

Note, however, that the travel-time change measured through the phase of an eigenray is not a measurement of phase velocity, as classically defined as the phase speed along the waveguide axis. For rays, the phase velocity along the waveguide axis is $c(z_0)/\cos(\theta_0)$, where $c(z_0)$ is the depth-dependent sound-speed profile at the source/receiver depth z_0, and θ_0 is the launch/receiver eigenray angle. In water, the bulk modulus shows nearly no frequency dependence, which means that wave dispersion in the ocean comes only from reverberation on interfaces and/or refraction due to sound-speed gradients. As a consequence, and with water being nondispersive, the group and phase velocities are the same along the ray path of an eigenray. The measurement of travel-time that changes through the phase of the DBF eigenray is then just a more accurate observable for the measuring of changes in the group velocity.

In the context of DBF, the sensitivity kernel is no longer point-to-point but relies on all source-receiver time series. The kernels are computed based on the concept that the processed signal is a linear combination of the time-delayed signals between all sources and receivers. Throughout this study, we concentrate on the physics that connects TSKs and Fresnel diffraction in the context of source-receiver arrays and a multipath environment in which DBF is performed to identify and isolate every eigenray. We show that TSKs associated with DBF result in increased spatial diversity of the sensitivity kernels which improves the range of validity of the ray theory for shallow-water acoustic tomography in the low-frequency regime.

This report is divided into four sections. Following this introduction (Sec. I), in Sec. II, the relationship between the TSK and the acoustic diffraction pattern is obtained for the point-to-point case. In Sec. III, the discussion is extended to source-receiver arrays through the DBF algorithm, which provides identification of every eigenray in the waveguide. The discussion continues in Sec. IV relating to the use of DBF in the context of shallow-water ocean acoustic tomography.

II. TSKs VERSUS DIFFRACTION

In this section, we investigate the links between TSKs and acoustic diffraction patterns for point-to-point, source-receiver configurations. Starting from the review of Skarsoulis *et al.*,[15] who first introduced TSKs into ocean acoustic tomography, we show here that in the far-field approximation, the TSK is the gradient of the diffraction pattern, corrected by a spatial factor.

The TSK is a measure of the travel-time perturbation of an acoustic path versus any spatial perturbation of the range-dependent and depth-dependent sound-speed profile. The pressure-field in the waveguide is expressed as the convolution of the source distribution over the source volume and the Green's function $G(\mathbf{r},\mathbf{r}_s,\omega)$. Under the first Born approximation, the Green's function perturbation δG has a linear relationship with the perturbation of the sound-speed distribution, δc, according to

$$\delta G(\mathbf{r}_r,\mathbf{r}_s,\omega) = -2\omega^2 \int\int\int G(\mathbf{r},\mathbf{r}_s,\omega) G(\mathbf{r}_r,\mathbf{r},\omega) \frac{\delta c(\mathbf{r})}{c^3(\mathbf{r})} dV(\mathbf{r}). \tag{1}$$

The temporal expression of the pressure-field $p(t)$ is written as the inverse Fourier transform of the frequency-domain pressure-field through

$$p(t) = \frac{1}{2\pi}\int G(\mathbf{r}_r,\mathbf{r}_s,\omega)P_s(\omega)e^{j\omega t}d\omega, \tag{2}$$

where $P_s(\omega)$ is the source spectrum. Then, a variation in the pressure-field δp has a linear relation with the Green's function perturbation:

$$\delta p(t) = \frac{1}{2\pi}\int \delta G(\mathbf{r}_r,\mathbf{r}_s,\omega)P_s(\omega)e^{j\omega t}d\omega. \tag{3}$$

Equations (3) and (1) provide linear relationship between the pressure-field perturbation and the sound-speed perturbation.

For estimation of the TSK, the perturbation of the travel-time related to the perturbation of the pressure-field needs to be considered. For acoustic propagation, Skarsoulis *et al.*[15] proposed that the travel-time is defined as the peak of the envelope of the demodulated or analytical signal. In the case of strong signal-to-noise ratios, as discussed above, the travel-time change is performed directly as the phase change of the pressure-field. In theory, this phase change can be measured at any time of the pressure-field. We chose to measure the travel-time change at the cycle peak of maximal amplitude τ_i. In this case, the relationship between the travel-time perturbation and the signal perturbation simplifies to

$$\delta\tau_i = -\frac{\delta\dot{p}_i}{\ddot{p}_i}, \tag{4}$$

where $\ddot{p}_i$ is the second-order derivative of the pressure-field p at time τ_i, and $\delta\dot{p}_i$ is the first-order derivative of the pressure-field perturbation δp evaluated at time τ_i [Eq. (3)]. Note that as compared to Skarsoulis *et al.*,[15] where they dealt with analytic complex signals, here the p and δp defined in Eqs. (2) and (3) are real quantities.

Combining Eqs. (1)–(4), the travel-time perturbation $\delta\tau_i$ is related to a perturbation of a sound-speed distribution δc through the integral:

$$\delta\tau_i = \int\int\int \delta c(\mathbf{r})K_i(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r)dV(\mathbf{r}), \tag{5}$$

where the expression K_i is the TSK. In Ref. 15, a general formulation of K_i is given for the case of analytic signals. Again, as we only deal with real signals, the TSK K_i simplifies to

$$K_i(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r) = \frac{1}{2\pi}\int \frac{j\omega}{\ddot{p}_i}Q(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r,\omega)e^{j\omega\tau_i}d\omega, \tag{6}$$

where Q is given according to[15]

$$Q(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r,\omega) = G(\mathbf{r},\mathbf{r}_s,\omega)G(\mathbf{r}_r,\mathbf{r},\omega)\frac{2\omega^2 Ps(\omega)}{c^3(\mathbf{r})}. \tag{7}$$

In Fig. 1, the TSK results are illustrated for two different geometries. The first was obtained through numerical simulation. For a given ray path in the waveguide, the TSK is built by computing Green's functions using a parabolic equation code[24] [Fig. 1(a)]. The second geometry was obtained through analytical Green's functions in a free-space medium where the receiver is the image of the actual receiver in the waveguide with respect to the waveguide boundary conditions [pressure release at the air-water interface and rigid bottom, Fig. 1(b)]. Consequently, the travel path in the free-space medium is the unfolded version of the actual travel path in the waveguide. The source signal has a 2.5 kHz central frequency and 4 kHz bandwidth. Interestingly, even if the acoustic fields are different in the two configurations, the two methods give similar TSK patterns away from the waveguide boundaries, where different echoes interfere [Fig. 1(c)].

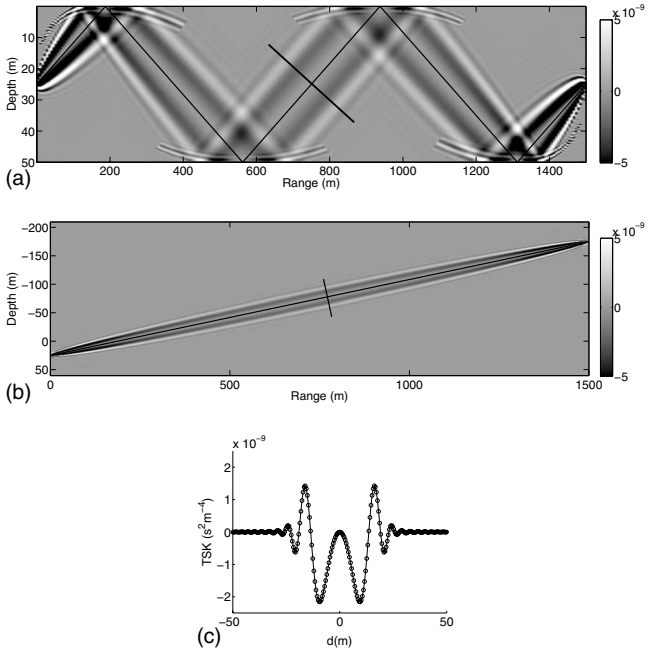


FIG. 1. (a) TSK (s^2 m^{-4}) in a waveguide for a ray path selected between two source-receiver arrays. (b) TSK for the equivalent ray path in free space [note scale change on vertical axis with respect to (a)]. (c) Cross-section of the TSK perpendicular to the ray path along the lines in (a) and (b) at the middle of the source-receiver range. Solid-line corresponds to (a), and circles correspond to (b).

The Green's function we consider in this waveguide is the sum of different eigenray contributions that can be separated through the DBF analysis. Each ray can have a complex trajectory (curved and/or broken lines), which can include rebounds on the waveguide boundaries. As we have been able to separate each ray from other neighboring rays, we can compute the travel-time and amplitude considering reflections at boundaries and/or any variations in speed. For simplicity, we will only consider here a uniform sound-speed and straight rays in the waveguide, although our study can also be extended to refracted rays. As stated above, the group and phase velocities along each eigenray are identical in this shallow-water regime, which means that the free-space approach will provide similar results away from the waveguide interfaces. Therefore, we proceeded with the free-space TSK for our analysis, and we can illustrate the results with the waveguide TSKs in some specific cases. Obviously, the computational costs in a free-space medium, where a simple analytic expression for the pressure-field is available, are much lower than in a waveguide where geometric dispersion has to be taken into account through modes or rays.

The standard well-known shape of the TSK is seen in Fig. 1, for the case of a uniform sound speed c. We observe zero-sensitivity on the ray path that refers to the so-called

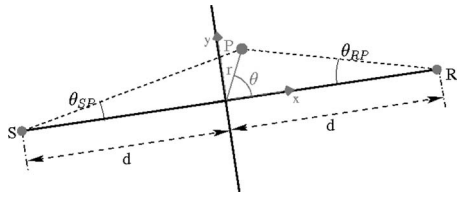


FIG. 2. Schematic of the source (s) and receiver (r) configuration in free space with polar coordinates (r,θ) and the definition of the angles θ_{RP} and θ_{SP}.

banana-doughnut shape of the TSK. A negative sensitivity zone is then seen, which is known as the first Fresnel zone. Higher-order Fresnel zones follow when moving away from the ray path.

On the other hand, the diffraction pattern between the source and the receiver has a maximum on the ray path. Its spatial derivative along a direction perpendicular to the ray is therefore zero on the ray. One may wish to investigate the link between this derivative and the TSK. Following the Huygens–Fresnel principle and invoking reciprocity, the acoustic diffraction pattern between a source and a receiver observed from any point of the medium $\mathbf{r}$ is computed as a product of both the Green's functions at the receiver $\mathbf{r}_r$ and the source $\mathbf{r}_s$:

$$D(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r,\omega)=G(\mathbf{r},\mathbf{r}_s,\omega)G(\mathbf{r},\mathbf{r}_r,\omega). \tag{8}$$

Considering the Green's function in a homogeneous free-space medium:

$$G(\mathbf{r},\mathbf{r}_0,\omega)=\frac{1}{4\pi d(\mathbf{r},\mathbf{r}_0)}e^{-j\omega d(\mathbf{r},\mathbf{r}_0)/c}, \tag{9}$$

it follows that

$$D(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r,\omega)=R(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r)e^{-j\omega\tau(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r)},$$

with

$$R(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r)=\frac{1}{(4\pi)^2 d(\mathbf{r},\mathbf{r}_s)d(\mathbf{r},\mathbf{r}_r)},$$

and

$$\tau(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r)=\frac{d(\mathbf{r},\mathbf{r}_s)+d(\mathbf{r},\mathbf{r}_r)}{c}. \tag{10}$$

Any change in $\mathbf{r}$ along the ray does not produce any phase change in the diffraction pattern D (since τ remains constant), and only produces small and smooth amplitude variations in the far field. On the contrary, a change in $\mathbf{r}$ along a perpendicular direction to the ray path will affect the phase, as investigated by Romanowicz and Snieder[25] and Snieder and Romanowicz[26] when considering seismic velocity perturbations. The gradient of the diffraction D on the perpendicular direction to the ray path is given by

$$\nabla_y D=\nabla D\cdot\boldsymbol{u}_y=[\nabla R-j\omega R\nabla\tau]e^{-j\omega\tau}\cdot\boldsymbol{u}_y, \tag{11}$$

where the $\boldsymbol{u}_y$ vector is the unitary vector along the y-direction shown in Fig. 2. At large distance from the source and receiver, ∇R can be ignored ($\nabla R\ll\omega R\nabla\tau$) and Eq. (11) reduces to

$$\nabla_y D=-j\omega R e^{-j\omega\tau}\nabla\tau\cdot\boldsymbol{u}_y. \tag{12}$$

Inside Eq. (12), there are (1) the propagation terms R and τ, which are related to both of the source/diffractor and receiver/diffractor propagations that determine the phase and amplitude of D; (2) the oblique gradient as the $\nabla\tau\cdot\boldsymbol{u}_y$ term, related to the local perturbations we are interested in. Considering the free-space Green's function G, we can rewrite Eq. (12) as

$$\nabla_y D=-j\omega G(\mathbf{r},\mathbf{r}_s,\omega)G(\mathbf{r},\mathbf{r}_r,\omega)(\nabla\tau\cdot\boldsymbol{u}_y), \tag{13}$$

Comparing Eqs. (7) and (13), we can see that the source-receiver diffraction pattern is proportional to the temporal derivative of Q multiplied by a spatial factor, $\nabla_y\tau$, which is analyzed below. Taking the reference point on the center of the ray-path trajectory, τ is written in polar coordinates (r,θ) as (see Fig. 2)

$$\tau(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r)=\frac{\sqrt{r^2+d^2-2rd\cos(\pi-\theta)}+\sqrt{r^2+d^2-2rd\cos(\theta)}}{c}. \tag{14}$$

Taking into account that $\cos(\pi-\theta)=-\cos(\theta)$ and $\sin(\pi-\theta)=\sin(\theta)$, the gradient of τ in polar coordinates becomes

$$\nabla\tau(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r)=\begin{pmatrix}\dfrac{r+d\cos\theta}{c\sqrt{r^2+d^2+2rd\cos(\theta)}}+\dfrac{r-d\cos\theta}{c\sqrt{r^2+d^2-2rd\cos(\theta)}}\\ \dfrac{-d\sin\theta}{c\sqrt{r^2+d^2+2rd\cos(\theta)}}+\dfrac{d\sin\theta}{c\sqrt{r^2+d^2-2rd\cos(\theta)}}\end{pmatrix}. \tag{15}$$

Calculating the directional gradient in the y-direction, we obtain

$$\nabla_y\tau=\nabla\tau\cdot\vec{u_y}=\frac{r\sin\theta}{c\sqrt{r^2+d^2+2rd\cos(\theta)}}+\frac{r\sin\theta}{c\sqrt{r^2+d^2-2rd\cos(\theta)}}=\frac{\sin\theta_{SP}+\sin\theta_{RP}}{c}, \tag{16}$$

where θ_{SP} and θ_{RP} are the source and receiver angles shown in Fig. 2. Finally, substituting Eq. (16) into Eq. (13) for the directional gradient of the diffraction pattern, and computing the inverse Fourier transform, we obtain the complete expression of the gradient of the diffraction pattern in the time domain as

$$\nabla_y D(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r)=-\frac{1}{2\pi}\int j\omega R(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r)e^{-j\omega\tau(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r)}\times\frac{\sin\theta_{SP}+\sin\theta_{RP}}{c}P_s(\omega)e^{j\omega\tau_i}d\omega, \tag{17}$$

which turns out to be

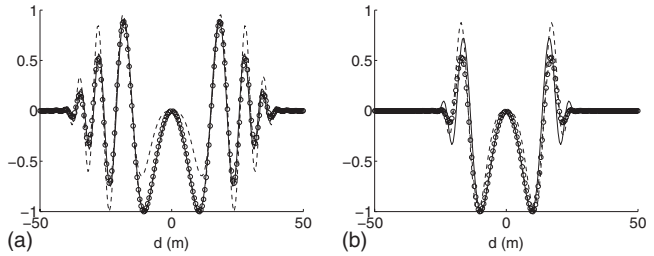


FIG. 3. [(a) and (b)] Point-to-point TSK (solid-line) versus diffraction pattern gradient (dashed) and spatially corrected diffraction pattern gradient (circle line). The three plots have been normalized according to their maxima. The source-receiver range is 1.5 km. The TSK was calculated at 750 m for a 7.6° acoustic ray for a 2.5 kHz central frequency pulse, and (a) 1250 Hz and (b) 4000 Hz frequency bandwidths.

$$\nabla_y D(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r) = -\frac{1}{2\pi}\frac{\sin\theta_{SP}+\sin\theta_{RP}}{c} \times \int j\omega G(\mathbf{r},\mathbf{r}_s,\omega)G(\mathbf{r},\mathbf{r}_r,\omega)P_s(\omega)e^{j\omega\tau_i}d\omega. \tag{18}$$

Similarly, combining Eqs. (6) and (7), we can write the TSK as

$$K_i(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r) = \frac{1}{2\pi}\int j\omega G(\mathbf{r},\mathbf{r}_s,\omega)G(\mathbf{r}_r,\mathbf{r},\omega)\frac{2\omega^2 P_s(\omega)}{\ddot{p}_i c^3(\mathbf{r})}e^{j\omega\tau_i}d\omega. \tag{19}$$

Equations (18) and (19) are similar but present two major differences. The first of these concerns the geometrical influence of the source/receiver position. Figure 3 shows the TSK (solid-line), the gradient of the diffraction pattern (dashed line), and the spatially corrected gradient of the diffraction pattern ($\nabla_y D(c/(\sin\theta_{SP}+\sin\theta_{RP}))$, circle line) for a 2.5 kHz central frequency signal with 1.25 and 4 kHz bandwidths, respectively. Figure 3 is obtained at a 750 m range position for a 1.5 km source-receiver range and 7.6° ray path. We can see in Fig. 3(a) that the use of the spatial correcting factor $(\sin\theta_{SP}+\sin\theta_{RP})/c$ allows a perfect fit between the TSK and the gradient of the diffraction pattern far from the waveguide interfaces. Note that this correction factor resembles the obliquity factor used in Fourier optics to account for diffraction effects from extended apertures.[27]

The second difference is more difficult to assess. The frequency contents of Eqs. (18) and (19) appear to be different, although Fig. 3(a) shows us that this is not the case. The frequency-dependent term $2\omega^2 Ps(\omega)/\ddot{p}_i c^3(\mathbf{r}')$ corresponds to the square of the acoustic wave number $(\omega/c)^2$ counterbalanced by the acceleration of the pressure-field. The acceleration of the pressure-field also justifies the missing minus term, as acceleration has an opposite phase with respect to pressure. As an illustration, Fig. 3(b) shows the same comparison performed over a larger bandwidth. We can see that the fit between Eqs. (18) and (19) is no longer perfect, since the ω^2 present in the TSK becomes significant when the bandwidth is large compared to the central frequency. We also note that the use of a wider bandwidth eliminates the side lobes of the TSK, putting sensitivity in the first Fresnel zone only. We will see below that the DBF processing leads to the same interesting phenomenon.

The similarity between Eqs. (18) and (19) is quite general and stays valid in more complicated waveguides with range-dependent and depth-dependent sound-speed patterns, as mentioned earlier. As long as we consider that the Green's function is a combination of well-identified separated eigenray contributions, we can isolate travel-time and amplitude for each eigenray connecting the source and the receiver [following Eqs. (9)–(13)], regardless of the path this ray has taken.

In conclusion, the point-to-point TSK is nothing else but the spatial derivative of the diffraction pattern that can be expected from the single-scattering (Born) approximation.

III. TSK VERSUS DIFFRACTION WITH DBF

When using DBF processes to extract eigenrays between two source-receiver arrays, the TSK reconstruction must take into account the geometry of the source and receiver arrays. More specifically, when using array processing, the measured travel-time does not correspond to the travel-time of a single arrival at one receiver $p(t)$ but corresponds to the travel-time of a linear combination of the properly time-delayed pressure-fields recorded between the source and receiver arrays. When performing the DBF analysis, the beam-formed pressure-field is given by

$$p_{\mathrm{BF}}(t,\theta_r,\theta_s) = \sum_{r=1}^{N_r}\sum_{s=1}^{N_s} a p_{rs}(t - T_r(\theta_r,z_r) - T_s(\theta_s,z_s)), \tag{20}$$

where $p_{rs}(t)$ is the pressure-field between source s and receiver r, and a_{rs} is an amplitude shading window. θ_r and θ_s are the observed receiver and launch angles, respectively, and T_r and T_s are the delay corrections to the beam-form on θ_r and θ_s. If the sound speed is uniform along the arrays, the time-delay beam-forming is

$$T(\theta,z) = (z - z_0)\frac{\sin\theta}{c}, \tag{21}$$

where $z-z_0$ is the distance between an array element at depth z and the center of the array at depth z_0.

The expression of an optimal delay correction for source-receive arrays in a depth-dependent sound-speed profile is given by the turning-point filter approach.[22] Physically speaking, the DBF consists of phasing a collection of signals according to given launch and receive angles; these are then averaged. After summation, the wave front associated with the launch and receiver angles is preserved, since it is coherently averaged, while the other field components are incoherently averaged and disappear. In this part, we investigate the spatial shape of the TSK when using DBF. As for the case of point-to-point TSK, two linearized expressions are required between δp_{BF} and δc, and between $\delta\tau_{\mathrm{BF}i}$ and δp_{BF}, to define the DBF-TSK.

As the DBF is a linear combination of pressure-fields between different source/receivers, we have the following expression:

$$\delta p_{\mathrm{BF}}(t,\theta_r,\theta_s)=\sum_{r=1}^{N_r}\sum_{s=1}^{N_s}a_{rs}\delta p_{rs}(t-T_r(\theta_r,z_r)-T_s(\theta_s,z_s)). \tag{22}$$

The linear relationship between δp and δc [see Eqs. (3) and (1)] implies a linear relationship between δp_{BF} and δc. Similarly, the relation between $\delta\tau_{\mathrm{BF}i}$ and δp_{BF} is given by Eq. (4), replacing p by p_{BF}.

Thus, using Eqs. (1), (3), (4), and (22), we define the linear relationship between the change in travel-time $\delta\tau_{\mathrm{BF}i}$ and the local perturbations of the sound-speed profile:

$$\delta\tau_{\mathrm{BF}i}=\int\int\int\delta c(\mathbf{r})K_{\mathrm{BF}i}(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r)dV(\mathbf{r}), \tag{23}$$

where the TSK is now:

$$K_{\mathrm{BF}i}(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r)=\frac{1}{2\pi}\int\frac{j\omega}{\ddot{p}_{\mathrm{BF}i}}Q_{\mathrm{BF}}(\mathbf{r},\mathbf{r}_{s0},\mathbf{r}_{r0},\omega)e^{j\omega\tau_i}d\omega, \tag{24}$$

with Q_{BF} as a linear combination of the point-to-point Q_{rs} expressed as

$$Q_{\mathrm{BF}}(\mathbf{r},\mathbf{r}_{s0},\mathbf{r}_{r0},\omega)=\sum_{r=1}^{N_r}\sum_{s=1}^{N_s}a_{rs}Q_{rs}(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r,\omega)e^{-j\omega(T_r(\theta_r)+T_s(\theta_s))}. \tag{25}$$

When using DBF, the diffraction pattern between the two arrays is given by

$$D_{\mathrm{BF}}(\mathbf{r},\mathbf{r}_{s0},\mathbf{r}_{r0},\omega)=\sum_{r=1}^{N_r}\sum_{s=1}^{N_s}a_{rs}D(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r,\omega)e^{-j\omega(T_r(\theta_r)+T_s(\theta_s))}. \tag{26}$$

If we take for D the expression in Eq. (10), and then compute the directional gradient of D_{BF} perpendicular to the ray path in the far-field approximation, we obtain the expression

$$\nabla_y D_{\mathrm{BF}}=-j\omega\sum_{r=1}^{N_r}\sum_{s=1}^{N_s}a_{rs}G(\mathbf{r},\mathbf{r}_s,\omega)G(\mathbf{r},\mathbf{r}_r,\omega)\times e^{-j\omega(T_r(\theta_r)+T_s(\theta_s))}(\nabla\tau(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r)\cdot\vec{u_y}). \tag{27}$$

The spatial factor $\nabla\tau(\mathbf{r},\mathbf{r}_s,\mathbf{r}_r)\cdot\vec{u_y}$ cannot be pulled out of the sum because of its specific dependence on each source and each receiver. So we cannot deduce the TSK from the spatial derivative of the diffraction pattern by a spatial correction factor, as we did previously in the point-to-point case. However, we can approximate this correction factor by the constant $\nabla\tau(\mathbf{r},\mathbf{r}_{s_0},\mathbf{r}_{r_0})\cdot\vec{u_y}$ at the central positions r_0 and s_0 of the source-receiver arrays. Actually, $\nabla\tau(\mathbf{r},\mathbf{r}_{s_0},\mathbf{r}_{r_0})\cdot\vec{u_y}$ also corresponds to the mean value of the directional gradients. Finally, we have the approximate relationship

$$K_{\mathrm{BF}i}(\mathbf{r},\mathbf{r}_{s0},\mathbf{r}_{r0})\approx\int\frac{\nabla_y D_{\mathrm{BF}}(\mathbf{r},\mathbf{r}_{s0},\mathbf{r}_{r0},\omega)}{\sin\theta_{S_0P}+\sin\theta_{R_0P}}cP_s(\omega)e^{-j\omega\tau_i}d\omega. \tag{28}$$

Figure 4 shows a comparison between $K_{\mathrm{BF}i}$ and the righthand term in Eq. (28). As expected, the comparison degrades when the size of the source-receiver arrays is increasing.

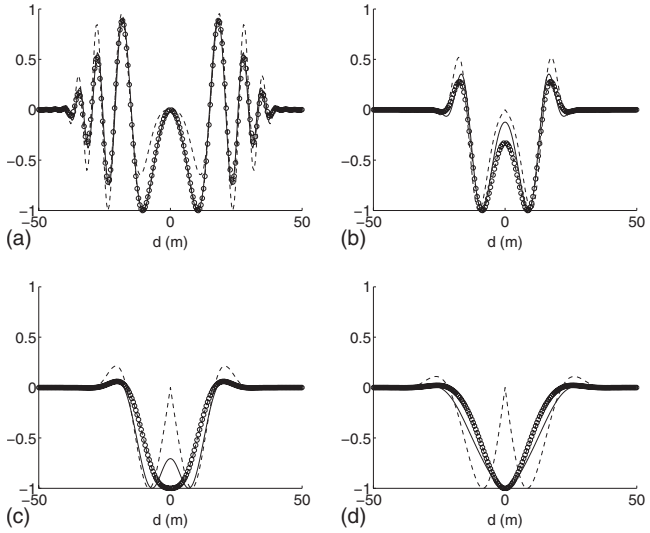


FIG. 4. (a) and (b) TSK versus the diffraction pattern gradient and spatially corrected diffraction pattern gradient, using the DBF with different array lengths. The three plots have been normalized according to their maxima. The source-receiver range is 1.5 km. TSK is calculated at 750 m for a 7.6° acoustic ray for a 2.5 kHz central frequency pulse and a 1250 Hz bandwidth. From (a) to (d), the source-receiver arrays are point to point, 30λ long, 60λ long, and 90λ long, respectively. The line codes are the same as in Fig. 3.

However, we can see that the correction factor $(\sin\theta_{SP}+\sin\theta_{RP})/c$ helps with a better approximation of the TSK from $\nabla_y D_{\mathrm{BF}}$, even in the case of the 90λ length array. In particular, we note that the directional gradient of the diffraction pattern has zero-sensitivity on the ray path, and that the spatial correction factor removes this zero-sensitivity when using arrays.

IV. ANALYSIS OF TSK USING DBF

The relationship between the TSK and the diffraction pattern of the source-receiver arrays is interesting since it can provide a simpler and faster way to calculate the TSK in shallow-water waveguides. Indeed, calculating TSK after DBF is a very time-consuming task, since it requires the computing of an ensemble of TSKs between each source-receiver pair, taken among the source-receiver arrays as required by Eqs. (24) and (25). Equation (28) shows that the TSK can be approximated from the product of the source and receiver array diffraction patterns only. For example, in the Fraunhofer approximation, the angle-dependent diffraction pattern of an N-element linear array around angle θ_0 at angular frequency ω is given by

$$B(\theta,\theta_0,\omega)=\frac{\sin(Na\omega(\sin\theta-\sin\theta_0)/2c)}{\sin(a\omega(\sin\theta-\sin\theta_0)/2c)}, \tag{29}$$

where a is the array pitch. In the far-field approximation, the diffraction pattern D_{BF} can be approximated by $B_s(\theta,\theta_s,\omega)B_r(\theta,\theta_r,\omega)$, where B_s and B_r are the diffraction patterns of the source and receiver arrays, respectively. Calculating the TSK $K_{\mathrm{BF}i}$ consists then in simply computing the directional gradient of the D_{BF} perpendicular to the ray path, weighted by the appropriate spatial correction factor, as shown in Eq. (28).

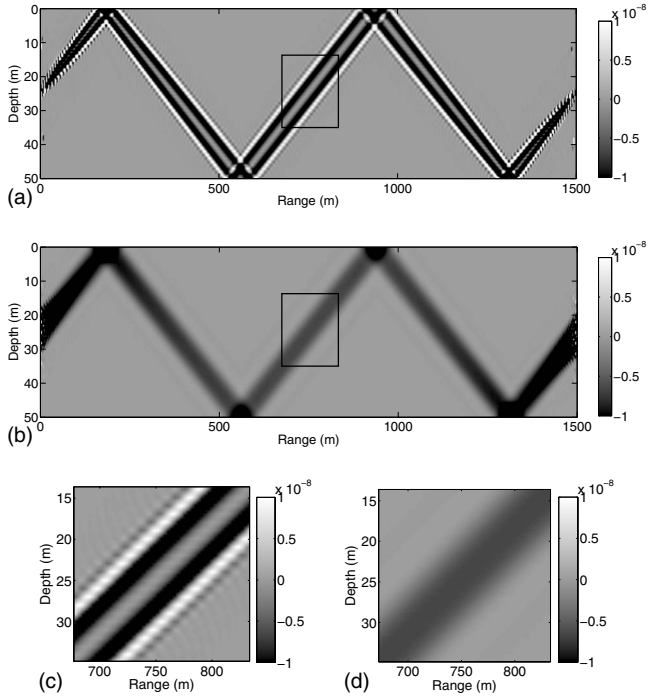


FIG. 5. TSKs ($s^2\ m^{-4}$) at 20 kHz central frequency with 30 kHz bandwidth. (a) point to point; (b) with DBF over 200−λ length arrays. [(c) and (d)] Zoom in on the central part of the ray, as defined by the black square in (a) and (b): (c) point to point; (d) with 200−λ length arrays.

Furthermore, Eq. (28) provides some physical insight in the shape evolution of the TSK with respect to the source-receiver array size. Two effects are revealed in Fig. 4. First, the side lobes become smaller when using extended antennas. This is an effect similar to the use of a larger frequency bandwidth [see the similarity between the solid-line in Fig. 3(b) and the solid-line in Fig. 4(b)]. Indeed, when integrating Eq. (29) over a frequency bandwidth Δf, it appears that an increase in Δf or array size Na leads to a similar decrease in secondary side lobes in the diffraction pattern. The same effect was seen by Raghukumar *et al.*[28] when analyzing the sensitivity kernel of a time-reversal mirror: they observed that the decrease in the sensitivity of the time-reversal focus to a sound-speed perturbation is due to the interference between more acoustic paths provided by the use of a larger array. The second effect is that the sensitivity on the ray path becomes nonzero and even becomes maximal with large aperture arrays (Fig. 4, solid-line). This result is non-intuitive, since it contradicts the well-known banana-doughnut shape classically observed with point-to-point TSK on the ray path. Indeed, as shown in Fig. 5(a), the point-to-point TSK remains zero on the ray path even at the high-frequency, large bandwidth limit. It is only the combination of TSK with DBF from source-receiver arrays that provides a non-zero-sensitivity of travel-times on the ray path. Indeed, the point-to-point zero-sensitivity on the ray path is a consequence of the stationary phase theorem that is associated to the Fermat principle that is no longer valid when the beam-forming process is performed on the source-receiver arrays.

In ocean acoustic tomography, such behavior for TSK is expected, assuming that the 3D TSK should be spatially averaged on two-dimensional (2D) sound-speed fluctuations in the ocean. In this case, TSK sensitivity on the ray path is obtained assuming a 2D TSK calculation for the sound-speed perturbation in the ocean. In our case, the maximal travel-time sensitivity on the ray path is due to the use of arrays on both sides of the waveguide and does not depend of assumptions on the shape of the ocean spatial fluctuations.

Of note, the high sensitivity seen after the DBF on the ray path is interesting from the point of view of ocean acoustic tomography. Indeed, in the limit of high-frequency, large bandwidth, and large arrays, the TSK in an acoustic waveguide shows a uniform kernel, with a sensitivity that is nearly limited to the ray path [Fig. 5(b)]. This further validates the use of ray theory and source-receiver arrays for shallow-water ocean acoustic tomography.

This last aspect of DBF and TSK leads to two interpretations. It could be said that if only a few rays are available in the waveguide, we may have more difficulties in locating anomalies, as the ambiguity of the perturbation position will now be on a broader zone. On the other hand, the uniform pattern of TSK after DBF in the Fresnel zone appears to have an advantage with respect to robustness. Indeed, sound-speed mismatch is always an issue with real data and will have a significant effect on the Fresnel zone oscillations seen on the point-to-point TSK shown in Fig. 5(c), while being negligible on the smoothly varying TSK measured after DBF [Fig. 5(d)].

V. CONCLUSIONS

In the present study, we have analyzed geometrically the mapping between TSKs and diffraction patterns. More precisely, we have shown that the point-to-point TSK is equivalent to the gradient of the diffraction pattern corrected by a spatial factor. When the pressure-field is double beam-formed on source-receiver arrays, we obtain an approximate relationship between the TSK and the diffraction pattern between the two arrays. Finally, the DBF process significantly modifies the spatial structure of TSK, in such a way that ocean acoustic tomography could improve the robustness of its performance when considering ray theory approximation.

ACKNOWLEDGMENTS

The authors would like to thank Bruce Cornuelle and Shane Walker for helpful comments on this study.

[1]P. Williamson, "A guide to the limits of resolution imposed by scattering in ray tomography," Geophysics **56**, 202–207 (1991).
[2]P. R. Williamson and M. H. Worthington, "Resolution limits in ray tomography due to wave behaviour: Numerical experiments," Geophysics **58**, 727–735 (1993).
[3]F. Dahlen, "Resolution limit of travel-time tomography," Geophys. J. Int. **157**, 315–331 (2004).
[4]J. Virieux and G. Lambaré, *Theory and Observation-Body Waves: Ray Methods and Finite Frequency Effects* (Elsevier, Amsterdam, 2007), pp. 127–155.
[5]S. Husen and E. Kissling, "Local earthquake tomography between rays and waves: Fat ray tomography," Phys. Earth Planet. Inter. **123**, 127–147 (2001).
[6]M. Woodward, "Wave-equation tomography," Geophysics **57**, 15–26 (1992).
[7]M. Born and E. Wolf, *Principles of Optics*, 6th ed. (Pergamon, New York, 1980).
[8]G. Nolet, *A Breviary of Seismic Tomography* (Cambridge University Press,

Cambridge, England, 2008).
[9]F. A. Dahlen, S. H. Hung, and G. Nolet, "Fréchet kernels for finite-difference travel-times—I. theory," Geophys. J. Int. **141**, 157–174 (2000).
[10]S. Gautier, G. Nolet, and J. Virieux, "Finite-frequency tomography in a crustal environment: Application to the western part of the Gulf of Corinth," Geophys. Prospect. **56**, 493–503 (2008).
[11]J. Tromp, C. Tape, and Q. Liu, "Seismic tomography, adjoint methods, time reversal, and banana-donut kernels," Geophys. J. Int. **160**, 195–216 (2005).
[12]M. V. de Hoop and R. D. van der Hilst, "On sensitivity kernels for 'wave-equation' transmission tomography," Geophys. J. Int. **160**, 621–633 (2005).
[13]A. Sieminski, J.-J. Lvque, and E. Debayle, "Can finite-frequency effects be accounted for in ray theory surface wave tomography?," Geophys. Res. Lett. **31**, L24614.1–L24614.4 (2004).
[14]P. Roux, W. Kuperman, W. Hodgkiss, H. C. Song, T. Akal, and M. Stevenson, "A non reciprocal implementation of time reversal in the ocean," J. Acoust. Soc. Am. **116**, 1009–1015 (2004).
[15]E. K. Skarsoulis and B. D. Cornuelle, "Travel-time sensitivity kernels in ocean acoustic tomography," J. Acoust. Soc. Am. **116**, 227–238 (2004).
[16]C. Tape, Q. Liu, and J. Tromp, "Finite frequency tomography using adjoint methods—Methodology and examples using membrane surface waves," Geophys. J. Int. **168**, 1105–1129 (2007).
[17]M. Fink, D. Cassereau, A. Derode, C. Prada, P. Roux, M. Tanter, J. Thomas, and F. Wu, "Time-reversed acoustics," Rep. Prog. Phys. **63**, 1933–1995 (2000).
[18]A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation* (Elsevier, Amsterdam, 1987).
[19]G. Herman, *Image Reconstruction From Projections: The Fundamentals of Computerized Tomography* (Academic, New York, 1980).
[20]A. J. Devaney, "Inverse-scattering theory within the Rytov approximation," Opt. Lett. **6**, 374–376 (1981).
[21]A. J. Devaney, "A filtered back-propagation algorithm for diffraction tomography," Ultrason. Imaging **4**, 336–350 (1982).
[22]M. Dzieciuch, P. Worcester, and W. Munk, "Turning point filters: Analysis of sound propagation on a gyre scale," J. Acoust. Soc. Am. **110**, 135–149 (2001).
[23]P. Roux, B. D. Cornuelle, W. A. Kuperman, and W. S. Hodgkiss, "The structure of raylike arrivals in a shallow-water waveguide," J. Acoust. Soc. Am. **124**, 3430–3439 (2008).
[24]M. Collins and E. Westwood, "A higher-order energy-conserving parabolic equation for range-dependent ocean depth, sound speed and density," J. Acoust. Soc. Am. **89**, 1068–1075 (1991).
[25]B. Romanowicz and R. Snieder, "A new formalism for the effect of lateral heterogeneity on normal modes and surface waves: II. General anisotropic perturbation," Geophys. J. Int. **93**, 91–99 (1998).
[26]R. Snieder and B. Romanowicz, "A new formalism for the effect of lateral heterogeneity on normal modes and surface waves: I. Isotropic perturbations, perturbations of interfaces and gravitational perturbations," Geophys. J. Int. **92**, 207–222 (1998).
[27]J. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, San Francisco, 1968).
[28]K. Raghukumar, B. D. Cornuelle, W. S. Hodgkiss, and W. A. Kuperman, "Pressure sensitivity kernels applied to time-reversal acoustics," J. Acoust. Soc. Am. **122**, 3023 (2007).

Patch near field acoustic holography based on particle velocity measurements

Yong-Bin Zhang[a)]
Institute of Sound and Vibration Research, Hefei University of Technology, Hefei 230009, China

Finn Jacobsen
Department of Electrical Engineering, Acoustic Technology, Technical University of Denmark, Building 352, Ørsteds Plads, DK-2800 Kongens Lyngby, Denmark

Chuan-Xing Bi and Xin-Zhao Chen
Institute of Sound and Vibration Research, Hefei University of Technology, Hefei 230009, China

(Received 3 April 2009; revised 21 April 2009; accepted 15 May 2009)

Patch near field acoustic holography (PNAH) based on sound pressure measurements makes it possible to reconstruct the source field near a source by measuring the sound pressure at positions on a surface that is comparable in size to the source region of concern. Particle velocity is an alternative input quantity for NAH, and the advantage of using the normal component of the particle velocity rather than the sound pressure as the input of conventional spatial Fourier transform based NAH and as the input of the statistically optimized variant of NAH has recently been demonstrated. This paper examines the use of particle velocity as the input of PNAH. Because the particle velocity decays faster toward the edges of the measurement aperture than the pressure does and because the wave number ratio that enters into the inverse propagator from pressure to velocity amplifies high spatial frequencies, PNAH based on particle velocity measurements can give better results than the pressure-based PNAH with a reduced number of iterations. A simulation study, as well as an experiment carried out with a pressure-velocity sound intensity probe, demonstrates these findings. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3158819]

PACS number(s): 43.60.Pt, 43.20.Ye, 43.40.At [EGW] Pages: 721–727

I. INTRODUCTION

Near field acoustic holography (NAH) is a useful technique for visualizing sound fields near complicated sources of sound. In addition to the classical NAH technique based on spatial discrete Fourier transforms,[1–3] many alternative methods have been developed in the past years, e.g., the inverse boundary element method,[4–6] the statistically optimized method,[7,8] the Helmholtz equation least-squares method,[9,10] and the equivalent source method.[11–13] Among all these techniques NAH based on spatial Fourier transforms is the easiest to implement and also the most computationally efficient because of the use of the fast Fourier transform algorithm. However, the spatial Fourier transform method requires that the measurement aperture covers a large region where the level drops to a sufficiently low value near the edges in order to reduce truncation effects. Often it is not possible to measure the sound field over such a large aperture, and even when it is possible it is likely to be very uneconomical. If the measurement aperture is not large enough, the reconstruction results will be significantly affected by the truncation effect; wrap-around errors and other errors due to the discontinuity of the aperture edge will be introduced.

In order to overcome this problem a technique called patch NAH (PNAH) has been proposed. In PNAH the sound field is only measured on a surface that is approximately as large as the patch on the source surface where the reconstruction is required. The key problem of PNAH is a numerical tangential extension of the measurement aperture. Saijyou and Yoshikawa[14] proposed a patch approach based on an iterative procedure, Williams and co-workers[15–17] extended it by using singular value decomposition and improved its accuracy by regularization, Sarkissian[18,19] developed a PNAH method using the superposition method, and Lee and Bolton[20] investigated PNAH in a cylindrical geometry. Yet another PNAH technique has been proposed by Steiner and Hald.[7]

All these techniques are based on measurement of the sound pressure. However, since a particle velocity transducer is available,[21] PNAH based on measurement of the normal component of the particle velocity is investigated in this paper. The use of the particle velocity in spatial Fourier transform based NAH has been examined by Jacobsen and Liu and found to be advantageous.[22] Statistically optimized NAH based on measurement of particle velocity has also been investigated, and the results showed a similar advantage.[23]

Compared with PNAH based on measurement of the sound pressure one might expect two advantages in using the normal component of the particle velocity. The normal component of the particle velocity generated by vibrations in a local source area decays faster toward the edges of the measurement aperture than the sound pressure does, and the particle velocity can therefore in general be expected to be "more local" than the pressure; thus the extrapolation of the

[a)]Author to whom correspondence should be addressed. Electronic mail: zybmy1997@163.com

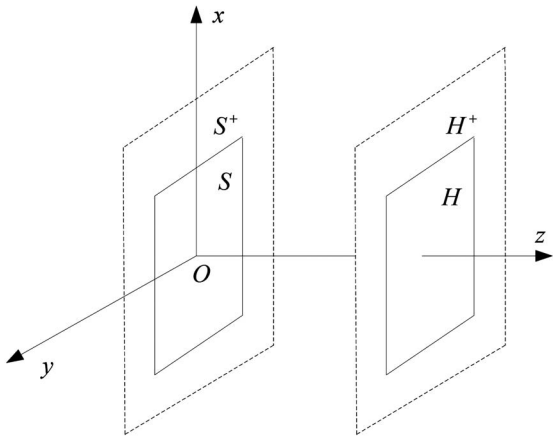


FIG. 1. Source and hologram plane.

sound field in measurement aperture might be expected to be more successful. Another potential advantage is associated with inherent properties of the inverse pressure-to-velocity propagator described in Ref. 22. The purpose of this paper is to examine the matter.

II. PNAH BASED ON MEASUREMENT OF PARTICLE VELOCITY

The derivation of the theory of PNAH based on measurement of the particle velocity follows the derivation for pressure-based PNAH in Ref. 15. Let the vector $\mathbf{v}_h$ represent the measured particle velocities at points $\mathbf{r}_h$ in a patch H, which is smaller than the sound source (see Fig. 1). The purpose is now to reconstruct the pressure and the normal component of the velocity at points $\mathbf{r}_s$ in a region S with the same size as H as accurately as possible. If the reconstruction is made directly using a spatial Fourier transform the result will be contaminated seriously by errors caused by the finite measurement aperture. These errors will be reduced if the measured particle velocity can be extrapolated into the region H^+. An iterative procedure for such a continuation of the measurement is described in Ref. 15. First the measured particle velocity is zero-padded,

$$\mathbf{v}_h^{(0)} = \begin{cases} \mathbf{v}_h & (\mathbf{r}_h \in H) \\ 0 & (\mathbf{r}_h \in H^+), \end{cases} \tag{1}$$

where $\mathbf{v}_h^0$ is the initial value of the iteration. The relationship between the filtered particle velocity $\tilde{\mathbf{v}}_h$ and the corresponding measured particle velocity for the ith iteration can be written as

$$\tilde{\mathbf{v}}_h^{(i)} = \mathbf{F}^{-1}\{F_\alpha^{(i)}\mathbf{F}\{\mathbf{v}_h^{(i)}\}\}, \tag{2}$$

where $\mathbf{F}$ and $\mathbf{F}^{-1}$ represent the discrete two-dimensional spatial Fourier transform and its inverse, and the filter factor $F_\alpha^{(i)}$ for the ith iteration is

$$F_\alpha^{(i)} = \frac{|G(k_x,k_y)|^2}{|G(k_x,k_y)|^2 + \alpha^{(i)}\left(\frac{\alpha^{(i)}}{\alpha^{(i)} + |G(k_x,k_y)|^2}\right)^2}. \tag{3}$$

In this equation $G(k_x,k_y)=e^{-jk_zd}$ is the propagator between the velocity of the surface of the source and the particle velocity in the hologram plane, d is the distance between the hologram plane and reconstruction plane (with the $e^{j\omega t}$ sign convention), the wave number component in the z-direction is given by

$$k_z = \begin{cases} \sqrt{k^2 - k_x^2 - k_y^2} & \text{for} \quad k_x^2 + k_y^2 \le k^2 \\ -j\sqrt{k_x^2 + k_y^2 - k^2} & \text{for} \quad k_x^2 + k_y^2 > k^2, \end{cases} \tag{4}$$

and $\alpha^{(i)}$ is the regularization parameter of the ith iteration, determined by the Morozov discrepancy principle[24] using

$$(\tilde{\mathbf{v}}_h^{(i)} - \mathbf{v}_h^{(i)})^H(\tilde{\mathbf{v}}_h^{(i)} - \mathbf{v}_h^{(i)})/M = (\sigma^{(i)})^2, \tag{5}$$

where M is the total number of points in the region $H+H^+$, and $\sigma^{(i)}$ is the standard deviation of the noise of the ith iteration, which is estimated as described in Ref. 24. In each iteration the regularization parameter can be determined by varying $\alpha^{(i)}$ until Eq. (5) is satisfied within a given tolerance. Here an alternative, fast method is used to determine $\alpha^{(i)}$ by searching for the minimum between the left part and the right part of Eq. (5) (using the function "fminbnd" of MATLAB).

The iteration runs as follows. Equations (1) and (2) give the filtered initial particle velocity. The input particle velocity for the first iteration is obtained by replacing the particle velocity in the region H with the original measured particle velocity $\mathbf{v}_h$, etc., that is,

$$\mathbf{v}_h^{(i+1)} = \begin{cases} \mathbf{v}_h & (\mathbf{r}_h \in H) \\ \tilde{\mathbf{v}}_h^{(i)} & (\mathbf{r}_h \in H^+). \end{cases} \tag{6}$$

The iteration can be stopped if the following condition is satisfied:

$$\frac{\|\tilde{\mathbf{v}}_h^{(i+1)} - \tilde{\mathbf{v}}_h^{(i)}\|_2}{\|\tilde{\mathbf{v}}_h^{(i)}\|_2} < \varepsilon, \tag{7}$$

where $\|\bullet\|_2$ represents the $L2$-norm of a vector and $\varepsilon > 0$ is a small number (for example, $\varepsilon = 10^{-5}$). Finally, the velocity of the source, $\mathbf{v}_s$, and the pressure on the surface of the source, $\mathbf{p}_s$, can be reconstructed from the filtered particle velocity and its continuation into the region H^+ after the iteration process has converged,

$$\mathbf{v}_s = \mathbf{F}^{-1}\{\mathbf{G}^{-1}F_\alpha^{(i)}\mathbf{F}\{\tilde{\mathbf{v}}_h^{(i)}\}\}, \tag{8}$$

$$\mathbf{p}_s = \mathbf{F}^{-1}\{\mathbf{G}_N^{-1}F_\alpha^{(i)}\mathbf{F}\{\tilde{\mathbf{v}}_h^{(i)}\}\}, \tag{9}$$

where $\mathbf{G}^{-1}(k_x,k_y)=e^{ik_zd}$ is the inverse velocity propagator, and

$$G_N^{-1} = \rho c(k/k_z)e^{jk_zd} \tag{10}$$

is the inverse of the propagator between the pressure in the source plane and the particle velocity in the hologram plane and ρc is the characteristic impedance of the medium. Note that the velocity-to-pressure propagator $\mathbf{G}_N^{-1}$ has a singularity near the radiation circle where k_z tends to zero. To reduce the singularity the propagator $\mathbf{G}_N^{-1}$ should be smoothed as described in Ref. 25.

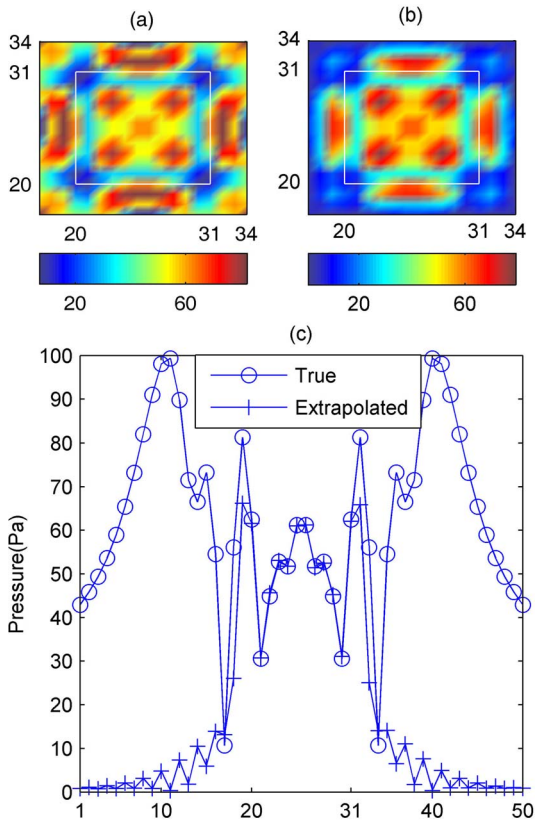


FIG. 2. (Color online) Extrapolation of the sound pressure at 2 kHz: (a) true pressure, (b) extrapolated pressure, and (c) comparison along a line. The frame in (a) and (b) indicates the boundary of the patch.

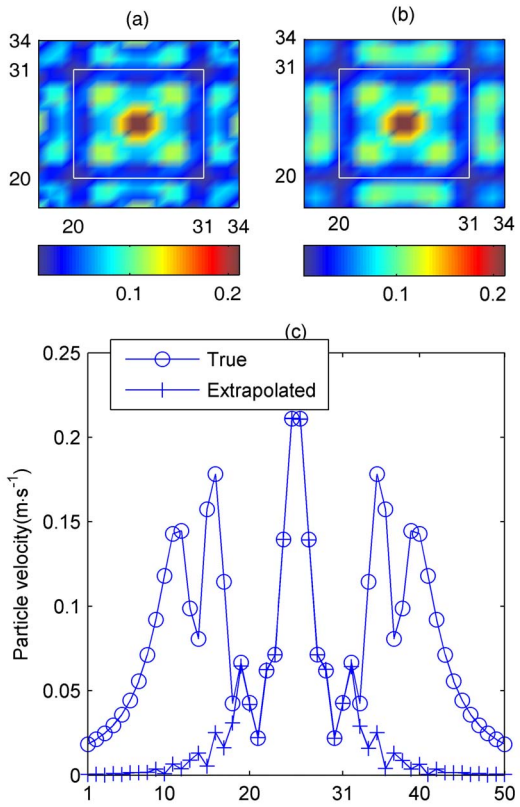


FIG. 3. (Color online) Extrapolation of the particle velocity at 2 kHz: (a) true velocity, (b) extrapolated velocity, and (c) comparison along a line. The frame in (a) and (b) indicates the boundary of the patch.

III. A SIMULATION STUDY

To examine the performance of the method described in the foregoing and compare it with the original version based on extrapolation of the sound pressure a simulation study has been carried out. The test case was a point driven simply supported 3-mm-thick aluminum plate in an infinite baffle. The dimensions of the plate were 0.58×0.58 m^2, and the excitation of the plate was a harmonic force with an amplitude of 100 N acting at the center of the plate. The reconstruction plane was located at $z=0.02$ m, and the hologram plane was located at $z=0.05$ m. The velocity of the surface of the plate was calculated by a modal sum, and the radiated sound field was calculated from a numerical approximation to Rayleigh's first integral. The sound pressure and the particle velocity were computed in the hologram plane in a grid 12×12 points with a lattice spacing in the x- and y-direction of 0.02 m. Thus, the measured data covered an area of 0.22×0.22 m^2, which is much smaller than the plate. The iteration procedure was applied both to the particle velocity and to the sound pressure in this patch in order to extrapolate the data to a region of 50×50 points with dimensions of 0.98×0.98 m^2. To make the simulation study more realistic noise has been added to the measured data corresponding to a signal-to-noise ratio of 35 dB. In the simulation the number ε in the convergence condition is 10^{-3}.

A. Extrapolation of sound field

Figures 2 and 3 compare the true and extrapolated sound pressure and particle velocity in the hologram plane at 2 kHz. Within the patch both the extrapolated sound pressure and particle velocity agree well with the true values. However, in the adjacent region outside the patch, the extrapolated pressure is in very poor agreement with the true pressure. The extrapolated particle velocity is more accurate than the extrapolated pressure, and, as can be seen in Fig. 3(b), reasonable agreement is obtained at the first two points outside the patch. However, it is clear that the region where the particle velocity can be extended accurately is limited. One may define the relative errors as

$$E_p = \frac{\|p_{\text{true}} - p_{\text{extrapolated}}\|_2}{\|p_{\text{true}}\|_2} \times 100\%, \tag{11a}$$

$$E_v = \frac{\|v_{h,\text{true}} - v_{h,\text{extrapolated}}\|_2}{\|v_{h,\text{true}}\|_2} \times 100\%. \tag{11b}$$

In a surrounding region one point, two points, and three points outside the patch the errors of the extrapolated sound pressure are 24.7%, 65.8%, and 91.3%, and the corresponding errors of the particle velocity are 17.4%, 32.8%, and 75.7%.

In addition to the increased reliability the extrapolation based on the particle velocity has another advantage, the iteration speed; for the same convergence condition, the re-

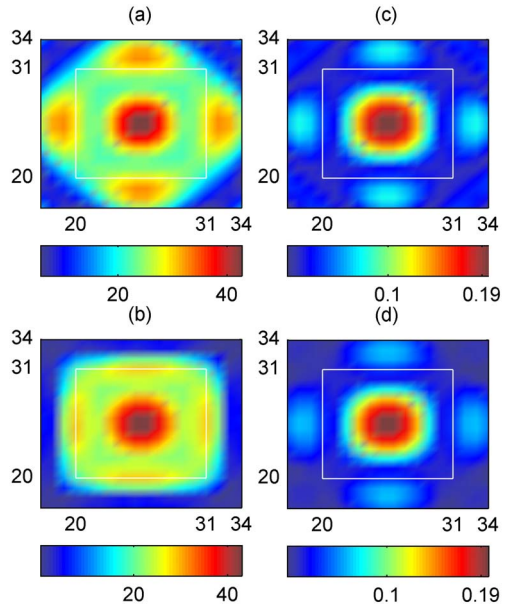


FIG. 4. (Color online) Extrapolation of sound pressure and particle velocity at 800 Hz: (a) true pressure, (b) extrapolated pressure, (c) true particle velocity, and (d) extrapolated particle velocity.

quired number of iterations for extrapolating the sound pressure was 415 in the example described above, whereas it was 203 for extrapolating the particle velocity.

It can be concluded that the particle velocity can lead to better extrapolation results than the pressure with less iterations. However, the number of required iterations for extrapolating the particle velocity is not always much smaller than the required number for extrapolating the pressure; for example, for the same test case at 800 Hz, the required numbers of iterations are 237 for the particle velocity and 271 for the pressure. Figure 4 compares the true and extrapolated sound pressure and particle velocity at 800 Hz. It can be seen that the extrapolated particle velocity is in better agreement with the true values than the extrapolated pressure is, although the particle velocity and the pressure require a similar number of iterations to converge. The errors defined as above are 31.8%, 85.1%, and 78.2% for the pressure and 14.1%, 26.0%, and 42.9% for the particle velocity. Other cases (not shown) have confirmed the tendency.

A better extrapolation of the pressure can be obtained if the number of the iterations is increased significantly. The extrapolated pressures at 2000 and 800 Hz after 1000 iterations are shown in Fig. 5. It can be seen that the accuracy of the extrapolation has been improved, while the computational cost and time have been increased.

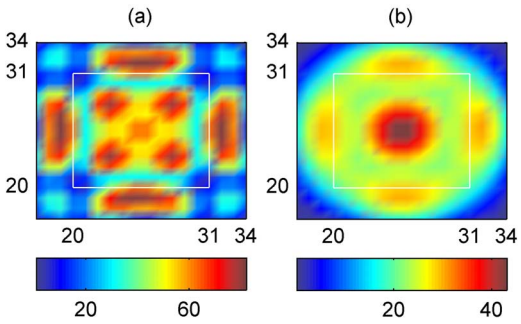


FIG. 5. (Color online) Extrapolation of the sound pressure at 2 kHz (a) and 800 Hz (b) after 1000 iterations.

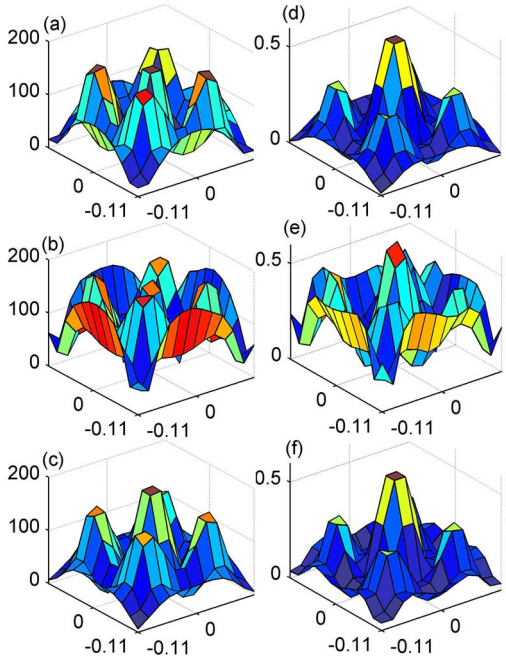


FIG. 6. (Color online) Reconstruction at 2 kHz: (a) true pressure, (b) pressure-to-pressure, (c) velocity-to-pressure, (d) true velocity, (e) pressure-to-velocity, and (f) velocity-to-velocity.

B. Reconstructed sound field

Figure 6 shows the reconstructed pressure and velocity in the source plane at 2 kHz. Parts (a), (b), and (c) compare

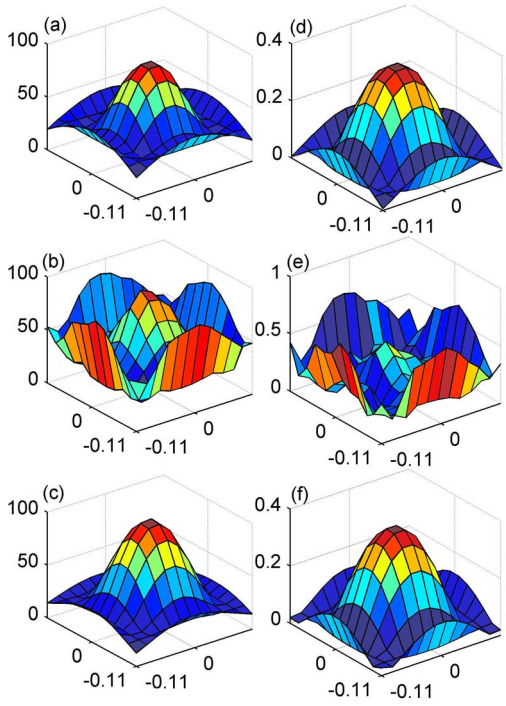


FIG. 7. (Color online) Reconstruction at 800 Hz: (a) true pressure, (b) pressure-to-pressure, (c) velocity-to-pressure, (d) true velocity, (e) pressure-to-velocity, and (f) velocity-to-velocity.

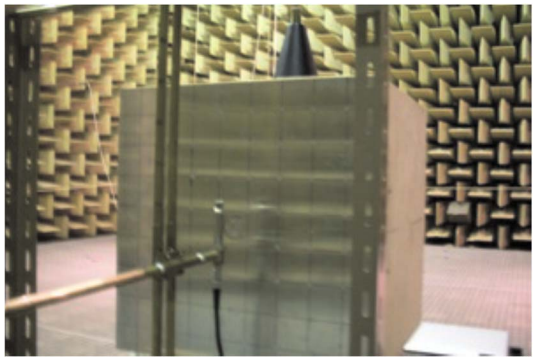


FIG. 8. (Color online) Experimental setup.

the true pressure, the pressure predicted from the extrapolated pressure, and the pressure predicted from the extrapolated particle velocity. It can be seen that the pressure reconstructed from the extrapolated particle velocity is in better agreement with the true pressure than the pressure reconstructed from pressure data, in particular, near the edges of the aperture. Parts (d), (e), and (f) compare the true particle velocity, the particle velocity predicted from the extrapolated pressure, and the particle velocity predicted from the extrapolated particle velocity. It is apparent that the particle velocity reconstructed from the pressure is not very accurate, whereas the reconstruction based on the particle velocity is far better.

Figure 7 shows a similar comparison when the plate is driven at 800 Hz. Again reconstructions based on the extrapolated particle velocity are fairly accurate, and, in particular, the velocity-to-velocity reconstruction is far better than the pressure-to-velocity reconstruction. By increasing the number of iterations for the pressure (as in Fig. 5) better reconstructions can be obtained at the expense of increased computer resources, and in some cases (not shown) the pressure-to-pressure reconstruction was found to be better than the velocity-to-pressure reconstruction, but velocity-to-velocity reconstructions were invariably found to be considerably better than pressure-to-velocity reconstructions.

IV. EXPERIMENTAL RESULTS

An experiment has been carried out in a large anechoic room at the Technical University of Denmark. The source was a 3-mm-thick aluminum plate with dimensions 44 $\times$44 cm^2 mounted as one of the surfaces of a box of heavy fiberboard and excited by a loudspeaker inside the box. The sound pressure and the particle velocity were measured at

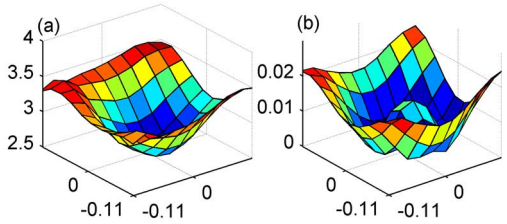


FIG. 9. (Color online) True (a) sound pressure and (b) particle velocity in the reconstruction plane.

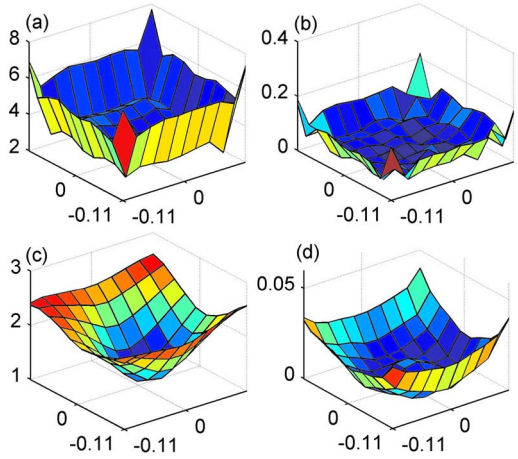


FIG. 10. (Color online) Reconstruction after 1500 iterations: (a) pressure-to-pressure, (b) pressure-to-velocity, (c) velocity-to-pressure, and (d) velocity-to-velocity.

18×19 points in two planes of dimensions 42.5×45 cm^2 using a $\frac{1}{2}$-in. *p-u* intensity probe produced by Microflown. The plate, the transducer, and the test fixture for the transducer are shown in Fig. 8. The transducer was calibrated as described in Ref. 26. The hologram plane was 8 cm from the plate, whereas the reconstruction plane was 4.5 cm from the plate and the measured data in it served as the "true" reference data. A Brüel & Kjær PULSE analyzer (type 3560) was used for measuring the frequency responses between the pressure and particle velocity signals from the transducer and the signal generated by the PULSE analyzer (pseudorandom noise) for driving the source.

A 10×10 point small patch with the center coincident with the center of the vibrating plate was selected for the extrapolation and reconstruction. Thus the patch covered an area of 22.5×22.5 cm^2. The number ε in the convergence condition was 10^{-5} in this experiment. Figure 9 shows the true pressure and particle velocity in the reconstruction plane at 160 Hz, Figs. 10(a) and 10(b) show the pressure and particle velocity predicted from the pressure in the measurement

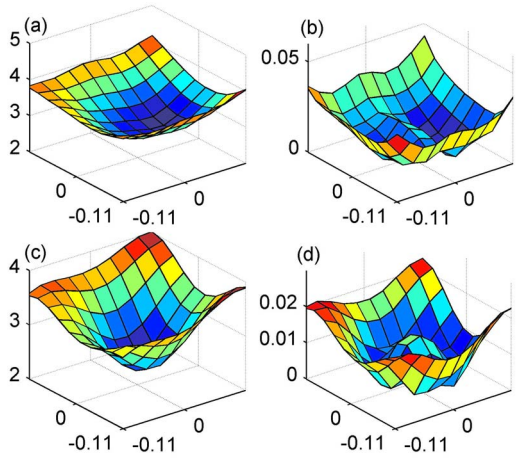


FIG. 11. (Color online) Reconstruction after convergence: (a) pressure-to-pressure, (b) pressure-to-velocity, (c) velocity-to-pressure, and (d) velocity-to-velocity.

TABLE I. Comparison of reconstructions based on pressure and on particle velocity.

Frequency (Hz)	Percent error of pressure reconstruction		Percent error of velocity reconstruction		Required number of iterations	
	Pressure-to-pressure	Velocity-to-pressure	Pressure-to-velocity	Velocity-to-velocity	Based on pressure	Based on velocity
100	10.5	8.4	54	7.1	14 744	4 295
200	6.5	7.9	25.5	4.2	19 282	16 697
400	7.2	24.4	43.2	13.6	26 518	23 027
800	10.0	22.3	35.1	10.4	15 500	13 153
1000	15.1	12.9	60.0	17.1	29 014	25 932

patch after 1500 iterations, and Figs. 10(c) and 10(d) show the pressure and particle velocity predicted from the particle velocity in the measurement patch after 1500 iterations. It can be seen that the reconstruction based on the extrapolated particle velocity is much better than the reconstruction based on the extrapolated pressure with the same number of iterations.

Figure 11 shows a similar comparison when convergence condition has been reached. It is apparent that the particle velocity performs better than the pressure as the input quantity. The relative error of the reconstructed pressure is 12.3% when it is based on pressure measurements and 6.13% when it is based on particle velocity measurements, and the corresponding error of the reconstructed particle velocity is 61.2% when it is based on pressure measurements and 4.33% when it is based on particle velocity measurements. It should also be mentioned that the extrapolated particle velocity converged after 20 790 iterations, whereas the extrapolated pressure required 28 142 iterations to converge. The results at some other frequencies are shown in Table I, which demonstrates that velocity-to-velocity reconstructions are generally better than pressure-to-velocity reconstructions. Sometimes pressure-to-pressure reconstructions are more accurate than velocity-to-pressure reconstructions. In all cases the iterative procedure converges faster for the particle velocity than for the pressure.

V. CONCLUSIONS

A method of PNAH based on extrapolating particle velocity data beyond the measurement aperture has been examined and shown to compare favorably with a similar method based on measurement of the sound pressure, both with respect to the number of required iterations and the quality of the resulting reconstructed sound field. These findings have been demonstrated by simulation as well as by an experiment carried out with a pressure-velocity intensity probe.

ACKNOWLEDGMENT

The authors would like to thank Microflown for lending the *p-u* sound intensity probe. This work was supported by the National Natural Science Foundation of China (Grant Nos. 50675056 and 10874037) and Fok Ying Tung Education Foundation (Grant No. 111058) and the Program for New Century Excellent Talents in University (Grant No. NCET-08–0767). Additionally, the China Scholarship Council is acknowledged for financial support.

[1] J. D. Maynard, E. G. Williams, and Y. Lee, "Nearfield acoustic holography: I. Theory of generalized holography and development of NAH," J. Acoust. Soc. Am. **78**, 1395–1413 (1985).
[2] W. A. Veronesi and J. D. Maynard, "Nearfield acoustic holography (NAH): II. Holographic reconstruction algorithms and computer implementation," J. Acoust. Soc. Am. **81**, 1307–1322 (1987).
[3] E. G. Williams, H. D. Dardy, and K. B. Washburn, "Generalized nearfield acoustical holography for cylindrical geometry: Theory and experiment," J. Acoust. Soc. Am. **81**, 389–407 (1987).
[4] M. R. Bai, "Application of BEM (boundary element method)-based acoustic holography to radiation analysis of sound sources with arbitrarily shaped geometries," J. Acoust. Soc. Am. **92**, 533–549 (1992).
[5] B.-K. Kim and J.-G. Ih, "On the reconstruction of the vibro-acoustic field over the surface enclosing an interior space using the boundary element method," J. Acoust. Soc. Am. **100**, 3003–3016 (1996).
[6] S.-C. Kang and J.-G. Ih, "Use of nonsingular boundary integral formulation for reducing errors due to near-field measurements in the boundary element method based near-field acoustic holography," J. Acoust. Soc. Am. **109**, 1320–1328 (2001).
[7] R. Steiner and J. Hald, "Near-field acoustical holography without the errors and limitations caused by the use of spatial DFT," Int. J. Acoust. Vib. **6**, 83–89 (2001).
[8] Y. T. Cho, J. S. Bolton, and J. Hald, "Source visualization by using statistically optimized nearfield acoustical holography in cylindrical coordinates," J. Acoust. Soc. Am. **118**, 2355–2364 (2005).
[9] Z. Wang and S. F. Wu, "Helmholtz equation-least-squares method for reconstructing the acoustic pressure field," J. Acoust. Soc. Am. **102**, 2020–2032 (1997).
[10] S. F. Wu and J. Yu, "Reconstructing interior acoustic pressure fields via Helmholtz equation least-squares method," J. Acoust. Soc. Am. **104**, 2054–2060 (1998).
[11] C.-X. Bi, X.-Z. Chen, and J. Chen, "Sound field separation technique based on equivalent source method and its application in nearfield acoustic holography," J. Acoust. Soc. Am. **123**, 1472–1478 (2008).
[12] C.-X. Bi, X.-Z. Chen, and J. Chen, "Nearfield acoustic holography based on the equivalent source method," Sci. China, Ser. E: Technol. Sci. **48**, 338–353 (2005).
[13] N. P. Valdivia and E. G. Williams, "Study on the comparison of the methods of equivalent sources and boundary element methods for near-field acoustic holography," J. Acoust. Soc. Am. **120**, 3694–3705 (2006).
[14] K. Saijyou and S. Yoshikawa, "Reduction methods of the reconstruction error for large-scale implementation of near-field acoustical holography," J. Acoust. Soc. Am. **110**, 2007–2023 (2001).
[15] E. G. Williams, B. H. Houston, and P. C. Herdic, "Fast Fourier transform and singular value decomposition formulations for patch nearfield acoustical holography," J. Acoust. Soc. Am. **114**, 1322–1333 (2003).
[16] K. Saijyou and H. Uchida, "Data extrapolation method for boundary element method-based near-field acoustical holography," J. Acoust. Soc. Am. **115**, 785–796 (2004).
[17] E. G. Williams, "Continuation of acoustic near-fields," J. Acoust. Soc. Am. **113**, 1273–1281 (2003).
[18] A. Sarkissian, "Extension of measurement surface in near-field acoustic holography," J. Acoust. Soc. Am. **115**, 1593–1596 (2004).
[19] A. Sarkissian, "Method of superposition applied to patch near-field acoustic holography," J. Acoust. Soc. Am. **118**, 671–678 (2005).
[20] M. Lee and J. S. Bolton, "Patch near-field acoustical holography in cylindrical geometry," J. Acoust. Soc. Am. **118**, 3721–3732 (2005).
[21] F. Jacobsen and H.-E. de Bree, "A comparison of two different sound intensity measurement principles," J. Acoust. Soc. Am. **118**, 1510–1517 (2005).

[22]F. Jacobsen and Y. Liu, "Near field acoustic holography with particle velocity transducers," J. Acoust. Soc. Am. **118**, 3139–3144 (2005).
[23]F. Jacobsen and V. Jaud, "Statistically optimized near field acoustic holography using an array of pressure-velocity probe," J. Acoust. Soc. Am. **121**, 1550–1558 (2007).
[24]E. G. Williams, "Regularization methods for near-field acoustical holography," J. Acoust. Soc. Am. **110**, 1976–1988 (2001).
[25]E. G. Williams and J. D. Maynard, "Numerical evaluation of the Rayleigh integral for planar radiators using the FFT," J. Acoust. Soc. Am. **72**, 2020–2030 (1982).
[26]F. Jacobsen and V. Jaud, "A note on the calibration of pressure-velocity sound intensity probes," J. Acoust. Soc. Am. **120**, 830–837 (2006).

Distortion-product otoacoustic emission input/output characteristics in normal-hearing and hearing-impaired human ears

Stephen T. Neely, Tiffany A. Johnson, Judy Kopun,
Darcia M. Dierking, and Michael P. Gorga
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 27 March 2009; revised 15 May 2009; accepted 26 May 2009)

Distortion-product otoacoustic emission (DPOAE) input/output (I/O) functions were measured in 322 ears of 176 subjects at as many as 8 f_2 frequencies per ear for a total of 1779 I/O functions. The f_2 frequencies ranged from 0.7 to 8 kHz in half-octave steps. Behavioral thresholds (BTs) at the f_2 frequencies ranged from −5 to 60 dB hearing loss (HL). Both linear-pressure and nonlinear, two-slope functions were fitted to the data. The two-slope function describes I/O compression as output-controlled self-suppression. Most I/O functions (96%) were better fitted by the two-slope method. DPOAE thresholds based on each method were used to predict BTs. Compared to estimates based on linear-pressure functions, individual BTs predicted from DPOAE thresholds based on the two-slope model had lower residual error and accounted for more variance. Another advantage of the two-slope method is that it provides an estimate of response growth rate (RGR) that is not tied to threshold. At all frequencies, the median low-level RGR (across I/O functions of the same f_2 and BT) usually increased as BT increased, while high-level compression decreased. The observed characteristics of DPOAE I/O functions are consistent with the loss of cochlear compression that is typically associated with mild-to-moderate HL. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3158859]

PACS number(s): 43.64.Jb [BLM] Pages: 728–738

I. INTRODUCTION

Distortion-product otoacoustic emissions (DPOAEs) are produced by the nonlinear interaction of two tones within the cochlea. This nonlinearity is due to the normal function of outer hair cells (OHCs), which is essential for normal auditory function, including auditory sensitivity. In fact, DPOAEs provide a window into cochlear function that allows estimates of behavioral threshold (BT) (e.g., Martin *et al.*, 1990; Boege and Janssen, 2002; Gorga *et al.*, 2003a). DPOAEs also provide information about the rate of growth of cochlear responses. Information about both of these features of cochlear function may be obtained in individual ears by quantifying characteristics of DPOAE input/output (I/O) functions.

Normal properties of DPOAEs have been studied extensively in humans and other animals (e.g., Probst *et al.*, 1991; Lonsbury-Martin and Martin, 2007), and share features that are at least qualitatively similar to those observed in invasive measurements of cochlear-response properties in lower animals. For example, I/O functions derived from direct basilar membrane (BM) measurements (Ruggero and Rich, 1991; Ruggero *et al.*, 1997; Rhode, 2007) are similar to DPOAE I/O functions (Mills *et al.*, 1993; Mills and Rubel, 1994). At the best or characteristic frequency (CF) for a particular location on the BM (defined as the frequency to which the place is most sensitive), there is nearly-linear growth in response to low stimulus levels, less-than-linear growth at moderate levels, and nearly linear growth again at high levels (e.g., Ruggero and Rich, 1991). However, when driven by a tone whose frequency is well below CF, growth rates remain nearly linear throughout the entire range of stimulus levels. Similar frequency dependence in slope has been observed when DPOAE responses are suppressed by a frequency that is about an octave below f_2, the higher frequency in the primary-frequency pair eliciting the DPOAE (Gorga *et al.*, 2003b, 2008).

Damage to OHCs is known to cause several changes to auditory function, including an increase in threshold (e.g., Dallos and Harris, 1978; Liberman and Dodds, 1984). Because the generation of DPOAEs depends on the integrity of the OHC system, DPOAEs are reduced or absent in ears with hearing loss (HL) (e.g., Kummer *et al.*, 1998; Gorga *et al.*, 1993, 2000; Dorn *et al.*, 2001). Since DPOAEs can be measured noninvasively, they have been used as a tool to detect impaired auditory function in humans (e.g., Lonsbury-Martin and Martin, 1990; Bonfils and Avan, 1992; Kim *et al.*, 1996; Gorga *et al.*, 1993, 1997). The clinical utility of these measurements is based on the knowledge that OHCs have essential roles in determining both DPOAE level and auditory sensitivity. In efforts to exploit this fact, some investigators have demonstrated a relationship between DPOAEs and BTs (Martin *et al.*, 1990; Gorga *et al.*, 1996; Boege and Janssen, 2002; Gorga *et al.*, 2003a; Rogers *et al.*, 2009).

Besides reducing auditory sensitivity, OHC damage changes the way cochlear responses grow with level. Following administration of furosemide, an agent known to reversibly reduce the endocochlear potential that serves as the power supply for OHC motility, auditory sensitivity is reduced and BM I/O functions at CF lose their characteristic compressive nonlinearity and become more linear (Ruggero

and Rich, 1991). Cochlear responses altered by furosemide are similar to response patterns observed with other more permanent cochlear impairments. Just as normal DPOAE I/O functions resemble directly-measured BM I/O functions in normal-hearing animals, the DPOAE I/O functions of hearing-impaired ears resemble BM I/O functions following cochlear insult (Mills *et al.*, 1993; Mills and Rubel, 1994; Dorn *et al.*, 2001).

In summary, data from lower animals reveal that direct measures of BM motion and indirect DPOAE measures of cochlear-response properties are similar and undergo similar changes in ears with induced cochlear lesions. While direct measurement of BM motion is impossible in humans, DPOAE measurements are noninvasive and are routinely used to detect HL. The similarities between direct cochlear-response measures in lower animals and indirect cochlear-response estimates based on DPOAE data (either in humans or lower animals) suggest that it may be possible to use DPOAE measurements in humans to provide insights into cochlear function over a wide range of levels.

The present study examines characteristics of DPOAE I/O functions in humans, with a goal of inferring cochlear-response properties for both normal and impaired ears. To the extent that DPOAE I/O functions are measures of cochlear-response growth, these comparisons help to describe the changes in response growth that occur as a consequence of HL in humans.

II. METHODS

DPOAE I/O functions were measured in 322 ears of 176 human subjects, ages 11–80 years old, at as many as eight f_2 frequencies per ear. Subjects were included if they demonstrated air-bone gaps <15 dB and had normal 226-Hz tympanograms just prior to when the DPOAE data were collected. The f_2 frequencies ranged from 0.7 to 8 kHz in half-octave steps. The level of f_2 (L_2) ranged from -20 to 80 dB sound pressure level (SPL) in 5-dB steps. The level of f_1 (L_1) and the ratio of primary frequencies (f_2/f_1) were determined by the following equations suggested by Johnson *et al.* (2006):

$$L_1 = 80 + 0.137 \log_2(18/f_2)(L_2 - 80), \quad (1)$$

$$f_2/f_1 = 1.22 + \log_2(9.6/f_2)(L_2/415)^2. \quad (2)$$

It was expected that for any given f_2 and L_2, the largest DPOAE level (on average for normal-hearing subjects) would be obtained when f_1 and L_1 are selected to satisfy Eqs. (1) and (2). However, obtaining the largest possible DPOAE level was not a requirement for this study.

At each f_2, I/O functions were sorted into 14 HL categories, ranging from -5 to 60 dB HL (see ANSI, 1996) in 5-dB steps, according to the BT of the ear at (or near) f_2.

A. Measurement

Subjects were seated in a sound-isolated room. Complete data collection required one or two 2-h sessions from each subject. Tympanograms were obtained at the beginning of each data-collection session. Stimuli were generated and responses recorded by a 24-bit soundcard (CardDeluxe, Digital Audio Laboratories) using locally-developed software (EMAV version 2.70; Neely and Liu, 1994). Sound was delivered to the ear by an insert earphone that included two separate sound sources and a calibrated microphone (ER-10C, Etymōtic Research). The ER-10C was modified to remove 20 dB of attenuation in order to allow presentation of stimuli at 80 dB SPL. Stimulus levels were calibrated in dB SPL in the ear using the ER-10C microphone. There are concerns for calibration errors when SPL is used (Siegel and Hirohata, 1994; Siegel, 2007; Scheperle *et al.*, 2008), especially at higher frequencies. However, recent data suggest that calibration method does not have a large impact on DPOAE test performance or the ability of DPOAEs to predict BT (Burke *et al.*, 2009; Rogers *et al.*, 2009).

For each f_2, data collection began at the highest L_2 and continued until the signal-to-noise ratio (SNR) <3 dB. DPOAE level (L_d) was based on the level in the $2f_1-f_2$ frequency bin, while noise level (L_n) included the level in the $2f_1-f_2$ bin as well as in five bins on either side of this bin. Levels were estimated by alternately storing 0.25-s samples in two buffers. To estimate L_d, the levels in the two buffers were summed and to estimate L_n, the levels in the two buffers were subtracted. Measurement-based stopping rules were used, in which data collection was terminated if the noise was <-25 dB SPL or artifact-free averaging time exceeded 32 s. The noise criterion, which places system distortion below the noise at all but the highest stimulus levels, was the primary stopping rule because its use resulted in the widest dynamic range of measurements across subjects and conditions. However, it was necessary to include an averaging-time criterion to avoid excessive averaging times for a single condition. In the majority of subjects, these rules allowed measuring DPOAEs for a wide range of levels, including low levels, and for a wide range of frequencies, including those at and below 1.5 kHz, where the noise floor sometimes is a problem.

B. Analysis

If complete I/O functions had been measured at each of 8 frequencies in every ear, a total of 2576 I/O functions would have been available for analysis. However, the inability to measure DPOAE levels above the noise floor in some subjects reduced the number of available I/O functions. This occurred mainly in subjects with HL, a number of whom did not produce DPOAEs at some frequencies, even for high-level stimuli. Additionally, in some subjects, it was not possible to reduce the noise levels to levels that would allow measuring responses reliably. This problem was frequency dependent, occurring almost exclusively for low f_2 frequencies. To be considered for further analysis, I/O functions were required to have a SNR of at least 3 dB at three or more consecutive stimulus levels. Although ears with greater HL were more likely to fail to meet this criterion, acceptable I/O functions were measured in some ears with thresholds as high as 60 dB HL. The distribution across frequencies and HL categories of the 1779 I/O functions considered acceptable for analysis is described in Table I.

TABLE I. The number of I/O functions in each HL category at each frequency that were included in the analysis.

	707	1000	1414	2000	2828	4000	5656	8000	Total
−5	2	6	7	5	6	18	12	12	68
0	25	20	36	27	26	37	28	18	217
5	54	53	35	31	37	21	28	29	288
10	32	42	36	36	21	17	19	22	225
15	33	31	22	23	22	14	16	15	176
20	21	21	18	14	16	6	10	12	118
25	19	19	25	11	5	10	8	6	103
30	7	11	12	15	4	8	7	6	70
35	3	11	13	19	19	14	10	3	92
40	2	7	18	16	13	15	8	11	90
45	7	5	7	11	12	13	15	9	79
50	5	4	7	13	14	17	18	15	93
55	1	4	8	8	13	15	17	11	77
60	1	1	1	5	16	29	12	18	83
Total	212	235	245	234	224	234	208	187	1779

Each individual I/O function was fitted by two different methods. In the first method, a linear regression of linear pressure (in μPa) onto L_2 (in dB SPL) was performed. This method was initially suggested by Boege and Janssen (2002) and has been used in other studies since its original description (e.g., Oswald and Janssen, 2003; Gorga *et al.*, 2003a; Rogers *et al.*, 2009).

In the second method, a nonlinear, two-slope function was fitted to the same data. The two-slope function models I/O compression as output-controlled suppression. The equation used to describe the two-slope function was motivated by the observation in previous studies of DPOAE suppression (Gorga *et al.*, 2003b, 2008) that decrements (amount of suppression) could be transformed into nearly-linear functions of suppressor level by the following equation:

$$D = 10 \log_{10}(10^{\text{decr}/10} - 1). \quad (3)$$

In this equation, decr represents the DPOAE decrement in decibel and D has been observed to have nearly-linear dependence on suppressor level. The success of this transformation in characterizing DPOAE suppression suggests that the decrement due to suppression is well described by

$$\text{decr} = 10 \log_{10}(10^{D/10} + 1), \quad (4)$$

where D is a linear function of suppressor level. Suppose that $D = S_2(L_s - L_i)$, where L_s represents suppressor level, L_i represents the input level (L_2) at the breakpoint between near-linear slopes of the I/O function at low levels and the more compressive slope at higher levels, and S_2 represents the amount of compression. S_2 and L_i are parameters of the linear dependence of D on suppressor level. With this substitution, the equation for decrement may be written as a function of suppressor level relative to the parameter L_i as follows:

$$\text{decr}(L_s - L_i) = 10 \log_{10}(10^{S_2(L_s - L_i)/10} + 1). \quad (5)$$

Our previous studies suggest that Eq. (5) provides a good description of DPOAE suppression (e.g., Gorga *et al.*, 2003b, 2008).

Suppose that compression of DPOAE I/O functions can be modeled as decrements (self-suppression) due to increases in DPOAE level. Equation (4) can be adapted to represent decrement in this model by substituting $D = S_2(L_d - L_o)$ as follows:

$$\text{decr}(L_d - L_o) = 10 \log_{10}(10^{S_2(L_d - L_o)/10} + 1). \quad (6)$$

In this equation, L_d represents DPOAE level and L_o represents the L_d at which the breakpoint between linear and compressive slopes occurred. The dependence of DPOAE level on stimulus level is presumed to be nearly linear when stimulus levels are low. The parameter S_1 represents the slope of the low-level portion of the I/O function. The nearly-linear and the compressive ranges are combined into a single, two-slope function that can be fitted to the DPOAE I/O functions

$$L_d = S_1(L_2 - L_i) + 10 \log_{10}(10^{S_2(L_d - L_o)/10} + 1). \quad (7)$$

Because L_d appears on both sides of Eq. (6), the fitting process is more conveniently performed by a nonlinear regression of L_d onto L_2 as follows:

$$L_2 = [L_d - 10 \log_{10}(10^{S_2(L_d - L_o)/10} + 1)]/S_1 + L_i. \quad (8)$$

Equation (8) has four parameters (L_i, L_o, S_1, and S_2), compared to only two parameters for the linear-pressure fit that has been used previously (i.e., Boege and Janssen, 2002). The increased number of parameters in the two-slope method is only justified if the resulting fits to the data are improved.

Prior to fitting I/O functions by either method, the DPOAE data were modified to force the I/O functions to have non-decreasing dependence on stimulus level. This requirement was imposed by replacing any DPOAE level that was less than the DPOAE level at the previous lower stimulus level with the larger DPOAE level. In effect, any local dips in the I/O functions were flattened out. Note that in both fitting methods, L_2 is the dependent variable of the regression. Flattening the I/O function avoids having disconnected L_2 values associated with any L_d. Elimination of the decreasing portions of the I/O function when fitting the data is jus-

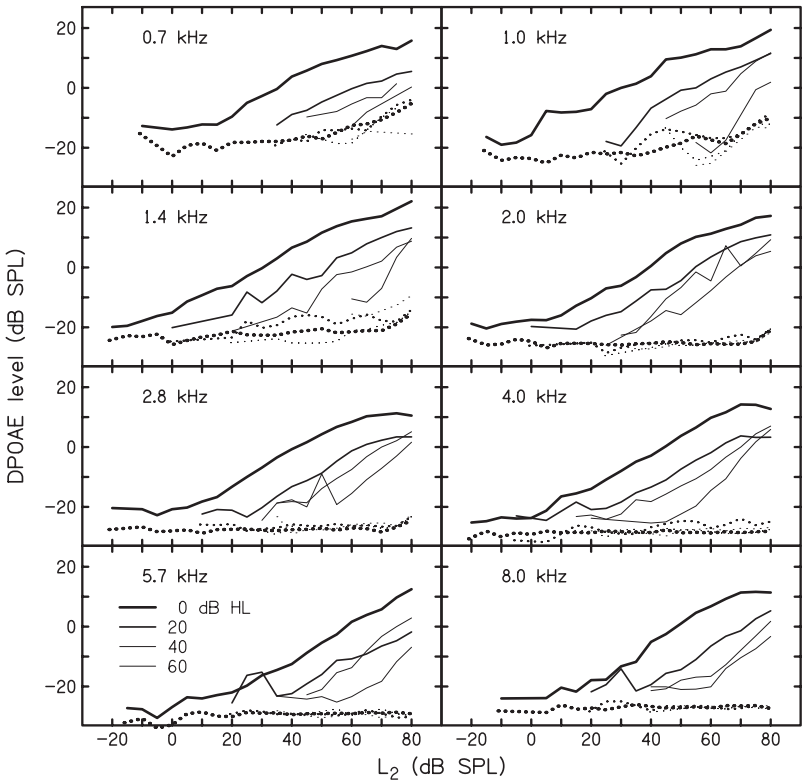


FIG. 1. Median DPOAE levels (solid lines) as a function of stimulus level. Each panel shows four different HL categories (0, 20, 40, and 60 dB HLs) at the indicated f_2 frequency. The HL category is indicated by the line weight from heavy (0 dB HL) to light (60 dB HL). The dashed lines indicate the median noise level at f_2 corresponding to each I/O function.

tified as a compromise required to allow the fitting of simple models to the data, such as the linear-pressure and two-slope models.

A nonlinear regression procedure was used to fit the two-slope function. In this procedure, the following limits were imposed to restrict the range of each parameter: $-20 < L_i < 80$, $-20 < L_o < 20$, $0.1 < S_1 < 1.5$, and $0.5 < S_2 < 8$. The importance of setting these parameter limits was in controlling behavior of the nonlinear regression for unusual cases.

DPOAE thresholds were estimated differently for the two fitting methods. For the linear-pressure fit, DPOAE threshold was defined as the intercept of the fitted function with zero pressure (e.g., Boege and Janssen, 2002). For the two-slope fit, DPOAE threshold was defined by setting $L_d = -25$ dB SPL in Eq. (8) and solving for L_2. DPOAE thresholds estimated by each method were used separately to predict BTs, and the predictive accuracy of the two methods was used as a quantitative approach to comparing the two fitting procedures. To improve the predictions from both methods, predicted BTs were restricted to the range from 0 to 60 dB HL.

III. RESULTS

Figure 1 shows how the shape of typical DPOAE I/O functions changes with frequency and HL category. These trends are illustrated by showing the median DPOAE level for representative HL categories (0, 20, 40, and 60 dB HLs) at each stimulus frequency (f_2). Across all HL categories and f_2 frequencies, I/O functions show increasing L_d with increasing stimulus level (L_2) with a maximum L_d of between 10 and 20 dB SPL. The modification to ensure nondecreasing I/O functions for fitting purposes was not done when obtaining median values for Fig. 1.

Each of the 1779 I/O functions was fitted by both linear-pressure and two-slope methods. The two-slope method provided a better fit for 96% of the I/O functions. A "better fit" was defined by one model resulting in a residual error that was at least 1% smaller than the other model. Of the remaining 4% that were not better fitted by the two-slope method, 1% were better fitted by the linear-pressure method and 3% were equally-well fitted by either method, meaning that the difference in the resulting residual error between the two methods was less than 1%. An examination of the distribution of better fits revealed no obvious trends for either the linear-pressure or two-slope methods across frequency and HL category.

Figure 2 shows an example of an I/O function that was better fitted by the two-slope method. In this case, the residual error was 74% less for the two-slope method. Note in the upper panel of Fig. 2 that DPOAE level is nearly a linear function of stimulus level over the entire range of stimulus levels. In other words, the I/O function has a nearly constant slope. The linear dependence of DPOAE level translates into exponential growth of pressure in the lower panel of Fig. 2, which was not well fitted by the linear-pressure model, as would be expected.

Figure 3 shows an example of an I/O function that was better fitted by the linear-pressure method. In this case, the residual error was 28% greater for the two-slope method. A visual comparison of the thin (linear-pressure) and thick (two-stage model) lines, however, reveals no apparent difference in the quality of the fits. Note how the change in slope of the two-slope function in the upper panel of Fig. 3 trans-

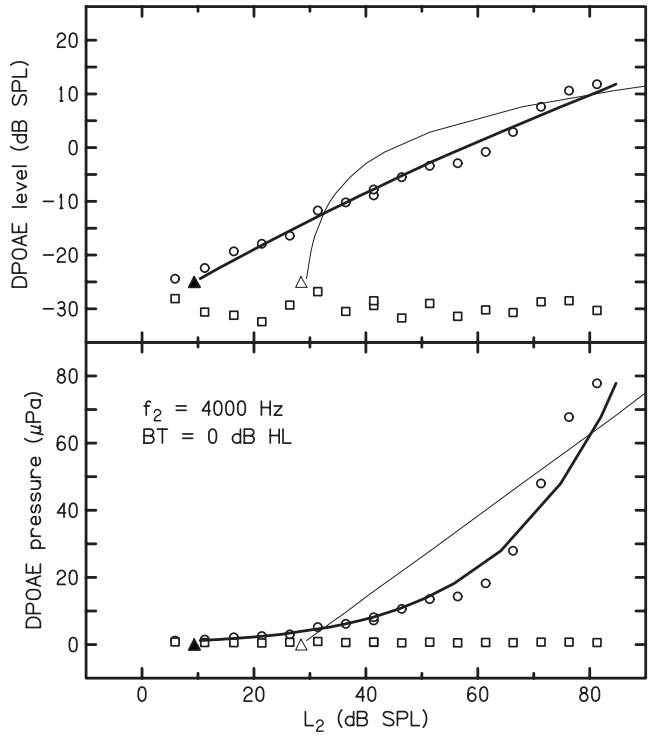


FIG. 2. Example of an I/O function better fitted by the two-slope model. The linear-pressure (thin line) and two-slope (thick line) fits are both superimposed on the DPOAE measurements at f_2=4 kHz in an ear with 0 dB HL threshold. The circles in the upper panel represent the measured DPOAE levels (L_d) and the circles in the lower panel represent the corresponding pressure amplitudes (P_d) in μPa. In this case, the standard error for the two-slope fit was 76% smaller than the linear-pressure fit. The squares represent the noise levels corresponding to each DPOAE measurement. The filled and open triangles indicate the DPOAE thresholds associated with the two-slope fit and linear-pressure fits, respectively.

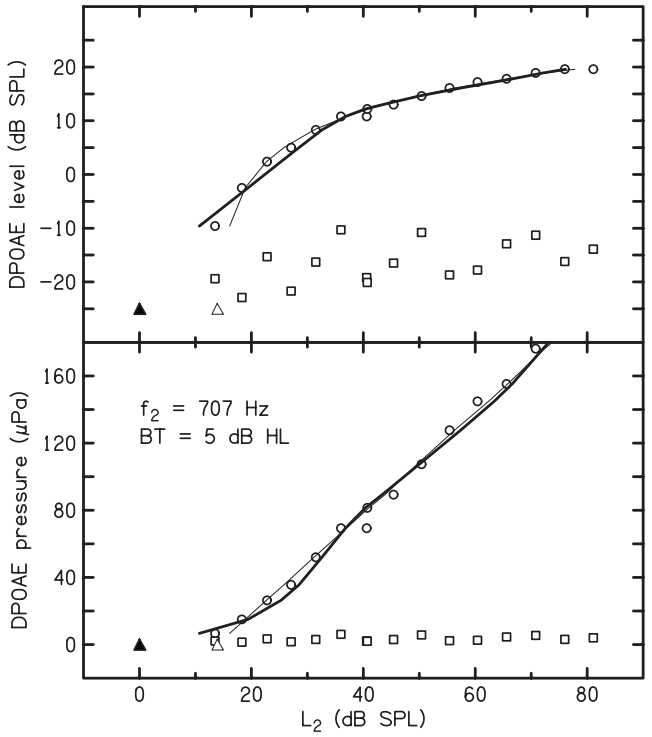


FIG. 3. Example of an I/O function better fitted by the linear-pressure model. The linear-pressure (thin line) and two-slope (thick line) fits are compared for a DPOAE I/O function at f_2=707 Hz in an ear with 5-dB HL threshold. The circles in the upper panel represent the measured DPOAE levels (L_d) and the circles in the lower panel represent the corresponding pressure amplitudes (P_d) in μPa. In this case, the standard error for the two-slope fit was 28% larger than the linear-pressure fit. The squares represent the noise levels corresponding to each DPOAE measurement. The filled and open triangles indicate the DPOAE thresholds associated with the two-slope fit and linear-pressure fits, respectively.

lates into a more linear function of pressure in the lower panel. This observation suggests that one reason for the success of the simpler linear-pressure method in fitting the DPOAE data is that many I/O functions are not well described by a single slope.

Figures 4–7 describe the level and slope parameters that were selected by the two-slope method to fit the DPOAE data. The median-level parameters for octave frequencies and each HL category are shown in Fig. 4. The level parameters describe the location of the breakpoint between the two slopes along the input and output axes of the I/O function in the top and bottom panels, respectively. The upper panel shows a trend of increasing L_i with BT, especially in the normal-hearing range from −5 to 20 dB HL. Increasing L_i means the breakpoint in the I/O function along the input-stimulus dimension moves toward the right. In other words, the onset of compression is lower in ears with better hearing. The lower panel shows L_o decreasing with BT, indicating that the breakpoint along the output-stimulus dimension tends to be lower with increasing HL. This result is expected because one would predict that the ear would produce less DPOAE output as BT increases.

Figure 5 shows the distribution of the level parameters within the 0 dB HL category for each f_2 frequency. The box-and-whisker plots show five characteristic points on these distributions. The top and bottom of the box represent the first and third quartiles, respectively, while the intermediate line bisecting the box represents the median. The upper and lower whiskers represent the furthest data points within 1.5 times the length of the box (i.e., 1.5 times the inter-quartile range) above the top or below the bottom of the box, respectively. In other words, the whiskers represent the furthest non-outlier data points. With two exceptions (0.7 and 5.6 kHz), the inter-quartile range of L_i (upper panel) is more than 10 dB, while the upper and lower whiskers can span nearly 40 dB. This large range of shifts in the input level (L_2) at which the break occurs is surprising considering that all of these ears had the same (0 dB HL) BTs. The distributions of L_o (lower panel) reveal that the range of vertical shifts in the output (L_d) at which the breakpoint occurs is similar to the range of horizontal shifts. There is no apparent trend in the median values of L_i or L_o across frequency; however, the median L_i appears to be skewed upward at 8 kHz, while the median L_o appears to be unusually low at 5.6 kHz. These anomalies will be discussed in more detail later.

Median-slope parameters are shown in Fig. 6. The lower-level slope S_1 (upper panel) tends to increase with BT, while S_2 tends to decrease. Note that median S_1 values are typically less than 1 when BT is less than 50 dB HL. Because S_2 multiplies L_d in Eq. (8), it may be interpreted as representing compression. This compression decreases the slope of the I/O function when $L_2 > L_o$. As a result, the high-

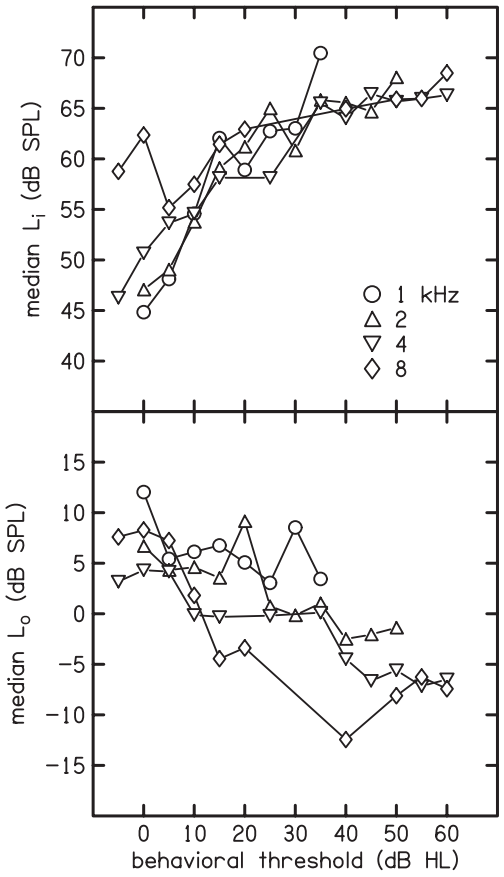


FIG. 4. Level parameters. The input-level parameter L_i (upper panel) represents the position of the two-slope breakpoint along the x axis, while the output-level parameter L_o (lower panel) represents the position of the two-slope breakpoint along the y axis.

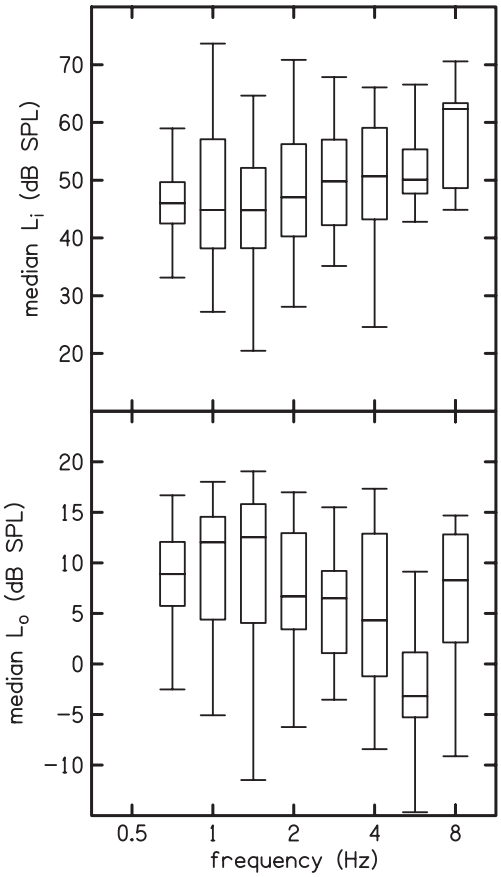


FIG. 5. Distribution of level parameters for 0 dB HL. Each box-and-whisker symbol indicates the median, inter-quartile range, and furthest non-outliers for selected distributions of either L_i (upper panel) or L_o (lower panel).

level slope of the two-slope function is S_1/S_2. The slope parameters in Fig. 6 reveal no obvious trend with frequency. Overall, the summary of median data in Fig. 6 indicates that the slope of the I/O function increases and the amount of compression decreases as threshold increases. Stated another way, response grows more rapidly when HL exists.

Figure 7 shows the distribution of the slope parameters within the 0 dB HL category for each f_2 frequency. Median values of S_1 are about 0.7 at all frequencies. Above 2 kHz, the inter-quartile range of S_1 is smaller, being only about 15% of its median value. The relatively small inter-quartile range suggests that S_1 is less variable across subjects than the other three parameters from the two-slope model. Median values of S_2 are less consistent across frequency compared to S_1. Note how the entire distribution of S_2 collapses to the minimum allowed value (0.5) at 5.6 kHz. This is the same frequency where the output-level parameter L_o is unusually low, which may indicate lower reverse middle-ear transmission at this frequency. Another interpretation is that the cochlea generates less distortion at 5.6 kHz, either due to a transition in cochlear mechanics or (perhaps more likely) to the particular choice of stimulus levels at this frequency.

Similar to L_i in Fig. 5, S_2 at 8 kHz is also skewed toward its maximum value in Fig. 7. These two trends reflect the fact that a large number of I/O functions at 8 kHz appear to become very compressed or saturate at a maximum DPOAE level when L_2 is large. In contrast, none of I/O functions at 5.6 kHz exhibited any compression at high levels, as evidenced by the collapse of the S_2 to its minimum value at this frequency in Fig. 7. Because stimulus parameters varied gradually across frequency, the contrast in observed I/O-function features between 5.6 and 8 kHz may indicate a transition in the underlying cochlear mechanics in this frequency region.

DPOAE thresholds (defined specifically and differently for each method as described above) are shown in Fig. 8 for both the linear-pressure (upper panel) and two-slope (lower panel) methods. For both methods, there appears to be a nearly-linear relation between DPOAE threshold and BT with no obvious difference in slope across frequency. The linear-pressure thresholds show less variability across frequency and have a shallower slope, compared to the two-slope thresholds. In practical terms, this means that DPOAE thresholds from the linear model were higher than BTs when BTs were less than 25 dB HL, compared to the two-slope model, but both models produced similar threshold estimates for greater HLs. If the dependent and independent variables were reversed in this figure (as might be the case in the clinic), then the two-slope model would provide threshold estimates in better agreement with BTs when hearing was within broadly defined normal limits. A linear regression of BT onto median DPOAE thresholds provided the coefficients needed to predict BT from DPOAE threshold. In this regres-

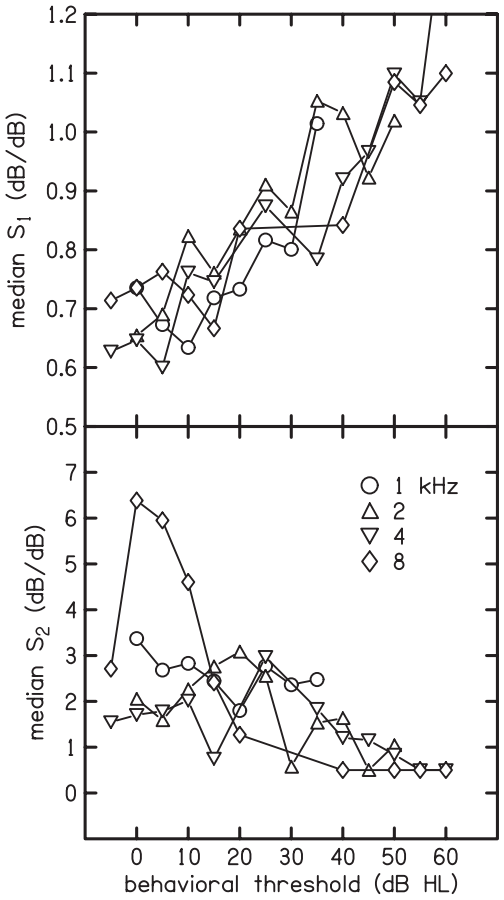


FIG. 6. Slope parameters. The first slope parameter S_1 (upper panel) represents the low-level slope, while the second parameter S_2 represents the additional compression that is applied to reduce the slope at high levels.

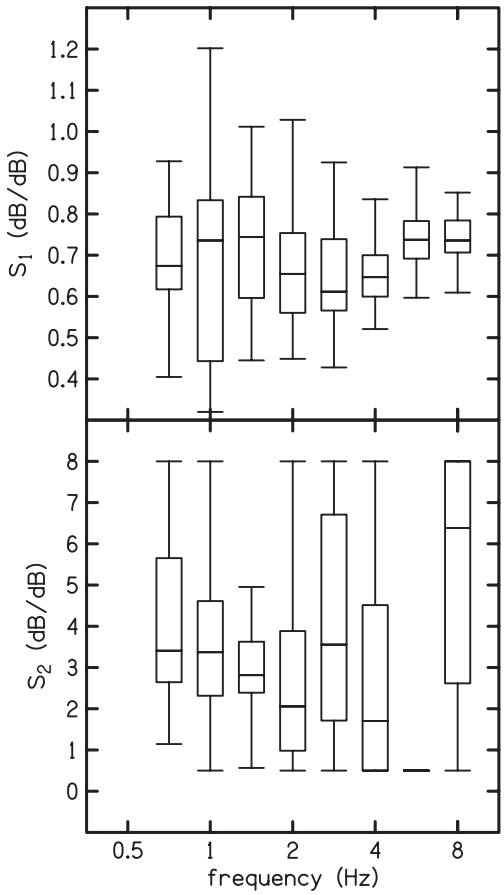


FIG. 7. Distribution of slope parameters for 0 dB HL. Each box-and-whisker symbol indicates the median, inter-quartile range, and furthest non-outliers for selected distributions of either S_1 (upper panel) or S_2 (lower panel).

sion, the coefficient that multiplies threshold was the same for all frequencies, but each frequency was allowed to have a different additive constant. The variance accounted for by this regression was essentially the same (96%) for both linear-pressure and two-slope methods. Adding all four two-slope parameters to the regression resulted in no improvement. The overall variance in individual BTs accounted for by the predicted thresholds was better for the two-slope method (57%) compared to the linear-pressure method (39%). Consistent with this improvement, the residual error in predicting BTs from DPOAE threshold was less with the two-slope method (12.5 dB) compared to the linear-pressure method (14.9 dB).

The distributions of residual errors across frequency and HL categories (shown in Fig. 9) reveal no obvious trends, except that the two-slope errors tend to be smaller when BTs are near zero, which follows from the summary provided in Fig. 8. The smaller prediction error at 0 dB HL may simply be due to the restriction on predicted BTs that prevented them from ever being less than zero, which was applied for both model predictions. The similarity between the two methods in the pattern of prediction errors suggests that these errors are predominantly due to variability inherent in I/O-function characteristics rather than being due to any failure to represent essential features of the I/O functions.

IV. DISCUSSION

Previous studies have used a variety of methods to model individual I/O functions. Boege and Janssen (2002) demonstrated that the linear-pressure fit produced DPOAE-threshold estimates that correlated with BT. Gorga *et al.* (2003a), using the linear-pressure method described by Boege and Janssen (2002), confirmed these results and suggested improvements by imposing limits on the ranges of both measured and predicted BTs. Oswald and Janssen (2003) suggested improvements in the linear-pressure threshold predictions by using weighted regressions that took into account the SNR of the measured DPOAE levels. Goldman *et al.* (2006) suggested an improvement to the linear-pressure method by applying a segmented-linear regression that incorporated more low-level responses into the regression, compared to Boege and Janssen (2002), giving it better detection performance. Compared to all of these previous methods, the two-slope method provides a better characterization of the rate of growth of DPOAE level over the entire range of stimulus levels. It is expected that the low-level slope parameter S_1 provides a useful measure of low-level DPOAE growth rate that may be compared with other measures of cochlear-response growth.

Decreasing dependence of DPOAE level with increasing stimulus level is more likely to occur in individual I/O func-

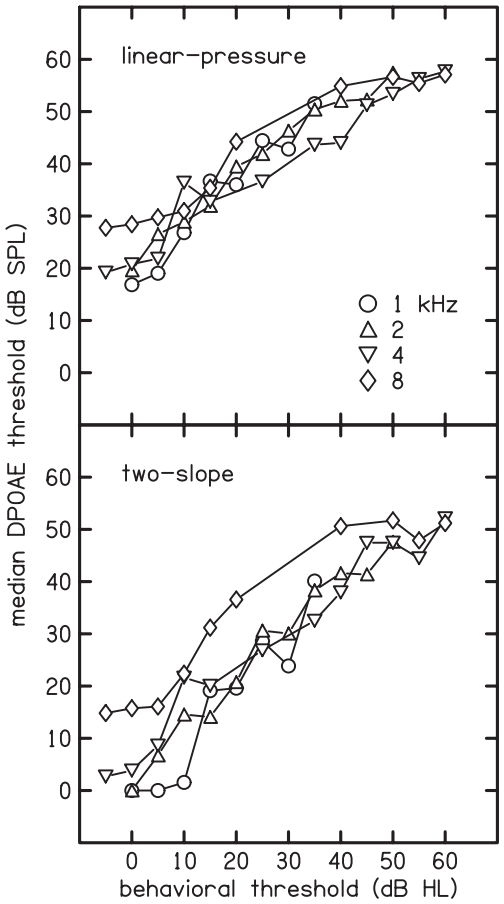


FIG. 8. DPOAE thresholds for both the linear-pressure (upper panel) and two-slope (lower panel) methods. For the linear-pressure method, threshold was defined as the intercept of the fitted line with the L_2 axis at zero pressure. For the two-slope method, threshold was defined as the L_2 value on the fitted function corresponding to L_d=−25 dB SPL. Examples of the relation of the DPOAE threshold to the fitted I/O function are illustrated by the triangles in Figs. 2 and 3.

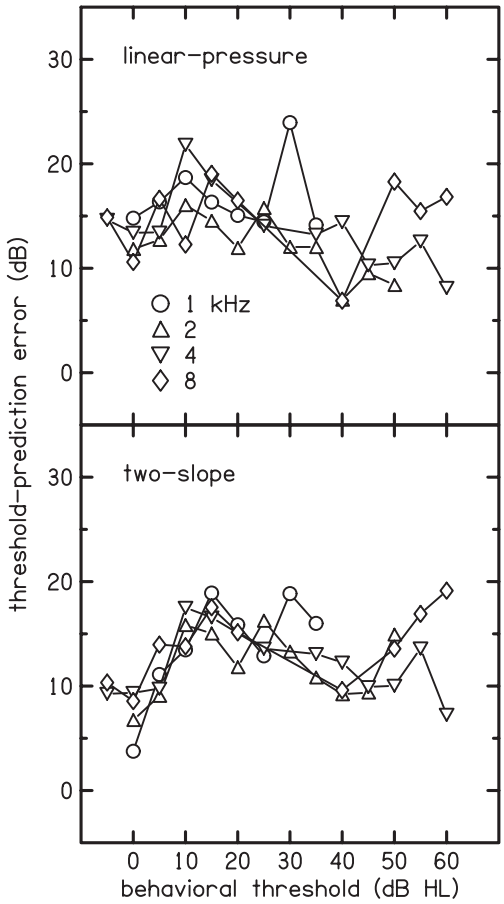


FIG. 9. Standard error of BT prediction based on DPOAE threshold for both the linear-pressure (upper panel) and two-slope (lower panel) methods. For both methods, the predicted BT was based on a linear regression of median DPOAE thresholds onto BT.

tions than in group averages. Such "dips" or "notches" in individual I/O functions do not occur consistently at particular levels, so they tend to be smoothed out by the averaging process. The presence of notches in I/O functions implies contributions from multiple sources within the cochlea that sometimes cancel each other due to having different phase. Notches were ignored when fitting individual I/O functions because of greater interest in characterizing response growth that might correlate with the perception of sound intensity than in the interaction of sources within the cochlea. Fitting notches in I/O functions would require more complex models than the linear-pressure and two-slope models that were employed in this study, and it is unclear as to whether the added complexity would improve the understanding of correlations between DPOAE response growth and the percept of intensity.

Fitting group-average DPOAE I/O functions is different than fitting individual I/O functions. Dorn *et al.* (2001) fitted an *ad-hoc* equation to median I/O functions from 77 subjects and observed some of the same trends with frequency, HL, and stimulus level as reported in the present study. Neely *et al.* (2003) noted the similarity in the shape of the average DPOAE I/O functions for normal-hearing ears with the shape of the loudness function described by Fletcher and Munson (1933). However, the logarithmic function that successfully characterized both DPOAE growth and loudness growth in that study failed to fit individual DPOAE I/O functions. Since only "average" functions of loudness growth were available in the Fletcher and Munson (1933) paper, it remains undetermined if the same problem exists for individual loudness functions as exists for individual DPOAE I/O functions.

Keefe (2002), Keefe and Abdala (2007), and Sanford *et al.* (2009) fitted rotated logistic functions to averaged DPOAE I/O functions from normal-hearing subjects. They suggested that the horizontal shift in the I/O function represents the influence of forward middle-ear transmission, while the vertical shift represents the influence of reverse middle-ear transmission. In a similar way, the level parameters of the two-slope function, which are plotted as functions of frequency in Fig. 10, may be interpreted as representing the influence of forward and reverse middle-ear transmissions. Any frequency dependence in forward middle-ear transmission should be observed in the frequency dependence of L_i and reverse transmission should be observed in L_o

The influence of forward or reverse middle-ear transmission on these input- and output-level parameters would be expected to be preserved across HL category, because the etiology of these HLs is thought to be within the cochlea and

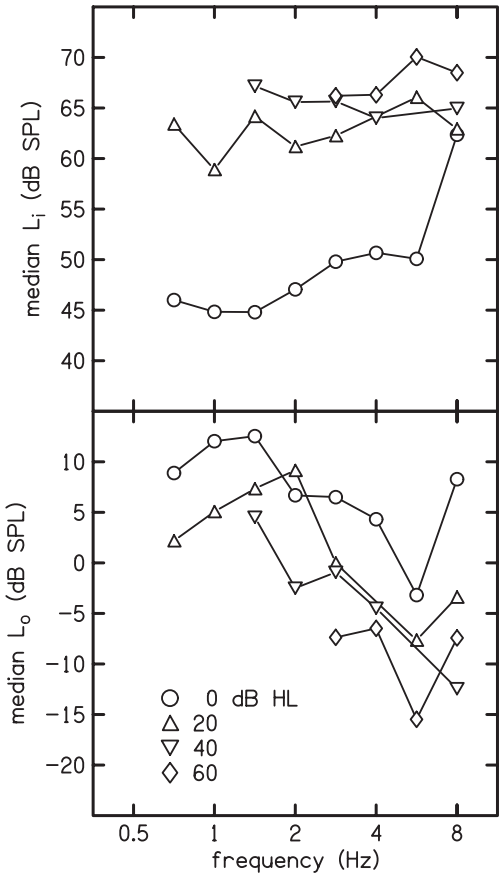


FIG. 10. Frequency dependence of two-slope level parameters for representative HL categories (0, 20, 40, and 60 dB HLs). The input-level parameter L_i (upper panel) includes the influence of forward middle-ear transmission, while the output-level parameter L_o (lower panel) includes the influence of reverse middle-ear transmission.

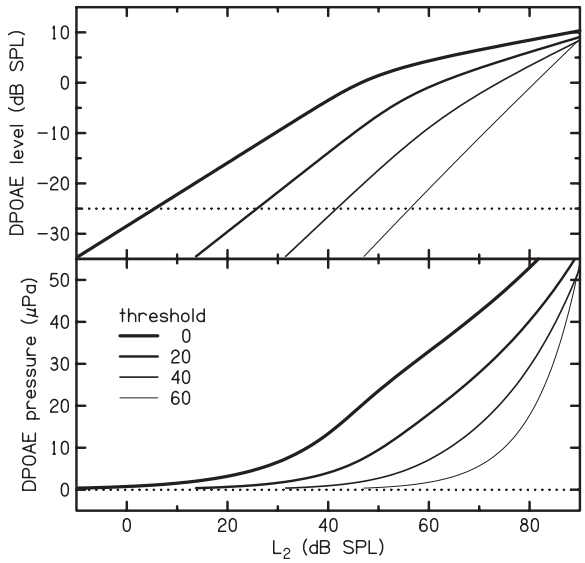


FIG. 11. Two-slope model I/O functions at 4 kHz for representative HL categories (0, 20, 40, and 60 dB HLs). DPOAE levels (dB SPL) are shown in the upper panel and the corresponding pressure (μPa) in the lower panel. These I/O functions were calculated using the regression coefficients listed in Table II.

not the middle ear, at least to the extent that a normal 226-Hz tympanogram just prior to DPOAE data collection indicates that the middle ear functions normally. The 6-dB rise in the input-level parameter L_i between 1 and 4 kHz (upper panel, Fig. 10) appears similar at 0 and 20 dB HL and may represent the influence of forward middle-ear transmission. However, a more obvious feature of L_i in Fig. 10 is the separation of 0 dB HL from the other HL categories. At most frequencies for the ears with BT=0 dB HL, the two-slope breakpoint typically occurs at input levels between 40 and 50 dB SPL, while the breakpoint moves above 60 dB SPL for ears with higher BTs. At f_2=8 kHz, the median L_i is above 60 dB SPL even for ears with thresholds of 0 dB HL. The output parameter L_o in the lower panel of Fig. 10 exhibits a notch at f_2=5.6 kHz for all HL categories. Because the output occurs at the $2f_1-f_2$ distortion frequency, a notch in L_o at f_2=5.6 kHz suggests that there maybe a notch in the reverse middle-ear transmission in the vicinity of 3.6 kHz.

Although the two-slope fit requires more computational effort than the linear-pressure fit and provided only minor improvement (as evaluated by the BT prediction error, which decreased from 14.9 to 12.5 dB), its major advantage may be in providing a meaningful quantification of I/O-function growth rate. Linear-regression coefficients of median two-slope parameters onto BT are listed in Table II. These coefficients allow trends in I/O-function shape to be visualized. For example, Fig. 11 demonstrates how model I/O functions at 4 kHz change as BT increases. Specifically, the breakpoint (between the two slopes) moves to the right and downward, while the low-level slope increases. Similar trends could be observed by using the regression coefficients for other frequencies. In the lower panel of Fig. 11, the same function that has two slopes when plotted in dB approximates a linear function over the range of pressures from 10 to 50 dB. This observation helps to explain why the linear-pressure fit performs as well as it does in predicting BTs.

The change in the shape with increasing HL category observed in the two-slope I/O functions shown in Fig. 11 is similar to the shape change observed in the median I/O functions in Fig. 1. However, few of the I/O functions in Fig. 1 show two distinct slopes. One reason why the median I/O functions lack the two-slope appearance may be that the in-

TABLE II. Linear-regression coefficients describing median I/O-function parameters as functions of BT at each frequency. As an example, the I/O functions based on these parameters at 4 kHz are shown in Fig. 11.

	C_{707}	C_{1000}	C_{1414}	C_{2000}	C_{2828}	C_{4000}	C_{5656}	C_{8000}	C_{BT}
L_i	53.4	52.4	51.1	50.1	51.9	50.3	53.0	54.7	0.3367
L_o	8.90	10.79	11.00	5.70	4.81	3.07	−2.01	1.48	−0.1798
S_1	0.602	0.689	0.716	0.698	0.622	0.627	0.708	0.650	0.0079
S_2	3.74	3.44	2.83	2.61	4.43	2.42	1.64	3.04	−0.0368

put level at which the breakpoint occurs differs across subjects within the same HL category, causing the breakpoint to be smoothed and less distinct in the median I/O functions. Individual differences in forward and/or reverse middle-ear transmission might also add variability in the input and output levels associated with the breakpoint. Group data would obscure the breakpoint under these conditions as well. Another reason that the breakpoint is less distinct in the group-average I/O functions may be that it is also less distinct in individual I/O functions. This appears to be the case in the example selected for Fig. 3, where there appears to be more curvature in the data below the breakpoint than in the fitted two-slope function. Despite the inability of the two-slope function to represent curvature below the breakpoint, it appears to provide a useful description of DPOAE I/O functions. The contribution to BT prediction error due to differences in the shape of the fitted and measured I/O functions appears to be small in comparison to other sources of variability.

The shape of DPOAE I/O functions for any particular L_2 and f_2 also depends on the selection of L_1 and f_1. In this study, the selection of L_1 and f_1 was based on the observations of Johnson *et al.* (2006), who described stimulus conditions (i.e., f_2-dependent f_2/f_1, and L_1 and L_2 combinations) thought to produce the largest L_d (on average) for normal-hearing subjects. If L_1 and f_1 had been chosen differently, the shapes of individual I/O functions may have been different. The similarity between the I/O-function shapes in Fig. 1 with those shown in Dorn *et al.*, 2001 (Fig. 3), where L_1 and f_1 were chosen differently, suggests that minor differences in the selection of L_1 and f_1 do not produce large differences in the I/O-function shapes.

It is possible that the growth rate of individual DPOAE I/O functions is correlated with loudness-growth rate in the same individuals, given the dependence of both measurements on the integrity of the cochlea. It is not known, however, whether S_1 or S_2 will exhibit more correlation with loudness-growth rate. Determining the extent of the correlation between loudness-growth rate and either S_1 or S_2 will require that loudness functions and DPOAE I/O functions both be measured in the same ears.

V. CONCLUSIONS

In addition to improving BT prediction, the two-slope function provides a description of individual DPOAE I/O functions that allows cochlear-response growth to be quantified. The low-level slope of the two-slope function is typically near 0.7 dB/dB in normal-hearing subjects and increases with HL. The variation of this low-level slope across subjects within the same HL category may represent individual differences in cochlear function, which could be compared to other measures of cochlear-response growth.

ACKNOWLEDGMENTS

The authors thank Liz Kennedy and Connie Converse for their contribution to data-collection efforts. This work is supported by the NIH/NIDCD (Grant Nos. R01 DC8318, R01 DC 2251, F32 DC7536, and P30 DC4662).

American National Standards Institute (**1996**). "Specifications for audiometers," ANSI S3.6-1996, New York.

Boege, P., and Janssen, T. (**2002**). "Pure-tone threshold estimation from extrapolated distortion product otoacoustic emission I/O-functions in normal and cochlear hearing loss ears," J. Acoust. Soc. Am. **111**, 1810–1818.

Bonfils, P., and Avan, P. (**1992**). "Distortion-product otoacoustic emissions. Values for clinical use," Arch. Otolaryngol. Head Neck Surg. **118**, 1069–1076.

Burke, S., Rogers, A. R., Neely, S. T., Kopun, J. G., Tan, H., and Gorga, M. P. (**2009**). "Influence of calibration method on DPOAE measurements: I. Test performance," presented at the American Auditory Society Meeting, Scottsdale, AZ.

Dallos, P. J., and Harris, D. M. (**1978**). "Two-tone rate suppression in auditory-nerve fibers: Dependence on suppressor frequency and level," Hear. Res. **49**, 225–246.

Dorn, P. A., Konrad-Martin, D., Neely, S. T., Keefe, D. H., and Gorga, M. P. (**2001**). "Distortion product otoacoustic emission input/output functions in normal and impaired ears," J. Acoust. Soc. Am. **110**, 3119–3131.

Fletcher, H., and Munson W. (**1933**). "Loudness: Its definition, measurement, and calculation," J. Acoust. Soc. Am. **5**, 82–108.

Goldman, B., Sheppard, L., Kujawa, S. G., and Seixas, N. S. (**2006**). "Modeling distortion product otoacoustic emission input/output functions using segmented regression," J. Acoust. Soc. Am. **120**, 2764–2776.

Gorga, M. P., Stover, L., Neely, S. T., and Montoya, D., (**1996**). "The use of cumulative distributions to determine critical values and levels of confidence for clinical distortion product otoacoustic emission measurements," J. Acoust. Soc. Am. **100**, 968–977.

Gorga, M. P., Neely, S. T., Bergman, B., Beauchaine, K. L., Kaminski, J. R., Peters, J., and Jesteadt, W. (**1993**). "Otoacoustic emissions from normal hearing and hearing-impaired subjects: Distortion product responses," J. Acoust. Soc. Am. **93**, 2050–2060.

Gorga, M. P., Neely, S. T., Dierking, D. M., Dorn, P. A., Hoover, B. M., and Fitzpatrick, D. (**2003b**). "Distortion product otoacoustic emission suppression tuning curves in normal-hearing and hearing-impaired human ears," J. Acoust. Soc. Am. **114**, 263–278.

Gorga, M. P., Neely, S. T., Dierking, D. M., Kopun, J., Jolkowski, K., Groenenboom, K., Tan, H., and Stiegemann, B. (**2008**). "Low-frequency and high-frequency distortion product otoacoustic emission suppression in humans," J. Acoust. Soc. Am. **123**, 2172–2190.

Gorga, M. P., Neely, S. T., Dorn, P. A., and Hoover, B. M. (**2003a**). "Further efforts to predict pure-tone thresholds from distortion product otoacoustic emission input/output functions," J. Acoust. Soc. Am. **113**, 3275–3284.

Gorga, M. P., Neely, S. T., Ohlrich, B., Hoover, B., Redner, J., and Peters, J. (**1997**). "From laboratory to clinic: A large scale study of distortion product otoacoustic emissions in ears with normal hearing and ears with hearing loss," Ear Hear. **18**, 440–455.

Gorga, M. P., Nelson, K., Davis, T., Dorn, P. A., and Neely, S. T. (**2000**). "Distortion product otoacoustic emission test performance when both $2f_1-f_2$ and $2f_2-f_1$ are used to predict auditory status," J. Acoust. Soc. Am. **107**, 2128–2135.

Johnson, T. A., Neely, S. T., Garner, C. A., and Gorga, M. P. (**2006**). "Influence of primary-level and primary-frequency ratios on human distortion product otoacoustic emissions," J. Acoust. Soc. Am. **119**, 418–428.

Keefe, D. H. (**2002**). "Spectral shapes of forward and reverse transfer functions between ear canal and cochlea estimated using DPOAE input/output functions," J. Acoust. Soc. Am. **111**, 249–260.

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.

Kim, D. O., Paparello, J., Jung, M. D., Smurzynski, J., and Sun, X. (**1996**). "Distortion product otoacoustic emission test of sensorineural hearing loss: Performance regarding sensitivity, specificity and receiver operating characteristics," Acta Oto-Laryngol. **116**, 3–11.

Kummer, P., Janssen, T., and Arnold, W. (**1998**). "The level and growth behavior of $2f_1-f_2$ distortion product otoacoustic emission and its relation to auditory sensitivity in normal hearing and cochlear hearing loss," J. Acoust. Soc. Am. **103**, 3431–3444.

Liberman, M. C., and Dodds, L. W. (**1984**). "Single-unit labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves," Hear. Res. **16**, 55–74.

Lonsbury-Martin, B. L., and Martin, G. K. (**1990**). "The clinical utility distortion-product otoacoustic emissions," Ear Hear. **11**, 144–154.

Lonsbury-Martin, B. L., and Martin, G. K. (**2007**). "Distortion-product otoacoustic emissions in populations with normal hearing sensitivity," in *Otoa-*

coustic Emissions: Clinical Applications, edited by M. S. Robinette and T. J. Glattke (Thieme, New York), pp. 107–130.
Martin, G. K., Probst, R., and Lonsbury-Martin, B. L. (**1990**). "Otoacoustic emissions in human ears: Normative findings," Ear Hear. **11**, 106–120.
Mills, D. M., Norton, S. J., and Rubel, E. W. (**1993**). "Vulnerability and adaptation of distortion product otoacoustic emissions to endocochlear potential variation," J. Acoust. Soc. Am. **94**, 2108–2122.
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 Liu, Z. (**1994**). "EMAV: Otoacoustic emission averager," Technical Memorandum 17, Boys Town National Research Hosptial, Omaha, NE.
Neely, S. T., Gorga, M. P., and Dorn, P. A. (**2003**). "Cochlear compression estimates from measurements of distortion-product otoacoustic emissions," J. Acoust. Soc. Am. **114**, 1499–1507.
Oswald, J. A., and Janssen, T. (**2003**). "Weighted DPOAE input/output functions: A tool for automatic assessment of hearing loss in clinical application," Z. Med. Phys. **13**, 93–98.
Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2067.
Rhode, W. S. (**2007**). "Basilar membrane mechanics in the 6–9 kHz region of sensitive chinchilla cochleae," J. Acoust. Soc. Am. **121**, 2792–2804.
Rogers, A. R., Burke, S., Neely, S. T., Kopun, J. G., Tan, H., and Gorga, M. P. (**2009**). "Influence of calibration method on DPOAE measurements: I. Threshold prediction," presented at the American Auditory Society Meeting, Scottsdale, AZ.
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. Neurol. **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.
Sanford, C. A., Keefe, D. H., Fitzpatrick, D. F., and Gorga, M. P. (**2009**). "Estimating forward and reverse ear-canal and middle-ear transmission in infants," presented at the American Auditory Society Meeting, Scottsdale, AZ.
Scheperle, R. A., Neely, S. T., Kopun, J. G., and Gorga, M. P. (**2008**). "Influence of *in-situ*, sound-level calibration on distortion-product otoacoustic emission variability," J. Acoust. Soc. Am. **124**, 288–300.
Siegel, J. H. (**2007**). "Calibrating otoacoustic emission probes," in *Otoacoustic Emissions: Clinical Applications*, edited by M. S. Robinette and T. J. Glattke (Thieme, New York), pp. 403–427.
Siegel, J. H., and Hirohata, E. T. (**1994**). "Sound calibration and distortion product otoacoustic emissions at high frequencies," Hear. Res. **80**, 146–152.

Limit cycle oscillations in a nonlinear state space model of the human cochlea

Emery M. Ku,[a] Stephen J. Elliott, and Ben Lineton
Institute of Sound and Vibration Research, University of Southampton, Southampton, Hampshire SO17 1BJ, United Kingdom

(Received 14 March 2009; revised 20 May 2009; accepted 26 May 2009)

It is somewhat surprising that linear analysis can account for so many features of the cochlea when it is inherently nonlinear. For example, the commonly detected spacing between adjacent spontaneous otoacoustic emissions (SOAEs) is often explained by a linear theory of "coherent reflection" [Zweig and Shera (1995). J. Acoust. Soc. Am. **98**, 2018–2047]. The nonlinear saturation of the cochlear amplifier is, however, believed to be responsible for stabilizing the amplitude of a SOAE. In this investigation, a state space model is used to first predict the linear instabilities that arise, given distributions of cochlear inhomogeneities, and then subsequently to simulate the time-varying spectra of the nonlinear models. By comparing nonlinear simulation results to linear predictions, it is demonstrated that nonlinear effects can have a strong impact on the steady-state response of an unstable cochlear model. Sharply tuned components that decay away exponentially within 100 ms are shown to be due to linearly resonant modes of the model generated by the cochlear inhomogeneities. Some oscillations at linearly unstable frequencies are suppressed over a longer time scale, whereas those that persist are due to linear instabilities and their distortion products. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3158861]

PACS number(s): 43.64.Kc, 43.64.Jb, 43.64.Bt, 43.40.Vn [BLM] Pages: 739–750

I. INTRODUCTION

Mammals rely upon an active mechanical process to enhance their sense of hearing. This takes place in the cochlea, a fluid-filled organ responsible for amplifying and converting the acoustically induced motion of its structures into neural impulses conveyed to the brain. A specialized structure in the human cochlea known as the organ of Corti (OC) performs this complex task. Within the OC are approximately 12 000 hair-like cells called outer hair cells (OHCs) (Pickles, 2003). These OHCs each contract and expand rapidly when a shearing motion opens and closes ion gates situated on the cells' stereocilia, thus inducing first mechano-electrical and then electro-mechanical transduction (Ashmore, 1987). The collective efforts of the OHCs comprise a system termed the cochlear amplifier (CA).

Experimental studies of the CA in animals have shown that it is capable of amplifying the magnitude of basilar membrane (BM) motion by up to ~45 dB at low excitation levels (Robles and Ruggero, 2001). At moderate levels, the CA begins to saturate and exhibit distortion effects; recent experimental work suggests that the distortion generated by the CA in mice may be due to the tip-links connecting adjacent stereocilia (Verpy *et al.*, 2008). At high levels, the output of the active CA is negligible relative to the passive excitation of the BM. The nonlinearity of the CA is thus an intrinsic feature of the normal cochlea that contributes to its enormous dynamic range, which is on the order of 120 dB sound pressure level (SPL) in humans (Gelfand, 1998).

One important group of cochlear epiphenomena is otoacoustic emissions (OAEs), sounds that propagate out of the cochlea and are detectable in the ear canal (Probst *et al.*, 1991). OAEs provide a non-invasive indication of cochlear health without the subject's participation and are in wide clinical use (Hall, 2000). Clinicians generally class OAEs by the stimulus used to evoke a response—such as clicks or tones. In trying to analyze how OAEs arise, Shera and Guinan (1999) proposed a classification paradigm based upon the generation mechanisms of OAEs—such as linear reflection, nonlinear distortion, or some combination thereof. One class of OAE that, at first glance, seems to fit neatly into both classification regimes is the spontaneous OAE (SOAE).

SOAEs are narrowband tones that propagate out of the cochlea without a stimulus and are detectable in the ear canal with a sensitive microphone (Probst *et al.*, 1986). Most human SOAEs are detected between 0.5 and 6 kHz, with the majority falling in the range of 1–2 kHz. Both the frequencies and amplitudes of SOAEs are remarkably stable over time (Probst *et al.*, 1991). These emissions are indicative of a healthy cochlea, being present in approximately 70% of all normal-hearing individuals (Talmadge and Tubis, 1993; Penner and Zhang, 1997). It is also well-known that the spectra of evoked emissions differ somewhat from those of SOAEs (e.g., Wable and Collet, 1994). So-called "long-lasting" evoked responses, also referred to as synchronized spontaneous otoacoustic emissions (SSOAEs), are sometimes recorded by extending the click evoked OAE (CEOAE) acquisition window from 20 to 80 ms (Gobsch and Tietze, 1993; Sisto *et al.*, 2001; Jedrzejczak *et al.*, 2008). The origin of this long-lasting ringing continues to be debated.

Early theoretical work by Gold (1948) that pre-dated the discovery of OAEs predicted the existence of SOAEs due to instability in locally active elements. However, after the discovery of OAEs by Kemp (1979), there were numerous reports of a commonly detected spacing between adjacent

[a]Author to whom correspondence should be addressed. Electronic mail: emk@soton.ac.uk

SOAE frequencies (Dallmayr, 1985, 1986; Talmadge and Tubis, 1993; Braun, 1997). Such regularity seemed to be at odds with a local oscillator framework, where one might assume unstable elements to be randomly distributed.

An alternative explanation of SOAE generation assumes multiple traveling wave (TW) reflections between a cochlear reflection site and the middle ear boundary at its base, analogous to standing waves in homogenous media (Kemp, 1979; Zwicker and Peisl, 1990; Zweig, 1991; Shera and Zweig, 1993; Talmadge and Tubis, 1993; Zweig and Shera, 1995; Allen *et al.*, 1995; Talmadge *et al.*, 1998; Shera and Guinan, 1999; Shera, 2003). Shera and Zweig (1993) and Zweig and Shera (1995) developed a linear "coherent reflection" hypothesis that built upon the cochlear standing wave theory and also accounted for the cause of the spacings between SOAEs.

Previous SOAE modeling work followed various avenues of research. Early reports showed that SOAEs can be suppressed and otherwise affected by external tones (e.g., Zwicker and Schloth, 1984). Phenomenological models using van der Pol oscillators were applied to describe the behavior of isolated SOAEs (e.g., Bialek and Wit, 1984; Wit, 1986; van Dijk and Wit, 1990; van Hengel and Maat, 1993; Murphy *et al.*, 1995). However, the success of the coherent-reflection theory in predicting many features of OAEs prompted modelers to try to replicate these findings in complete mechanical models of the cochlea. Including random perturbations in the smoothly varying mechanical parameters was found to generate reflections and cochlear standing waves in a variety of models (e.g., Talmadge *et al.*, 1998; Talmadge *et al.*, 2000; Shera *et al.*, 2005; Elliott *et al.*, 2007).

The nonlinear time domain simulations of Talmadge *et al.* (1998) showed amplitude-stabilized limit cycle oscillations (LCOs) similar to SOAEs, but the number, frequencies, and amplitudes were not predicted. In contrast, the linear stability tests performed by Ku *et al.* (2008) predicted the number and frequencies of instabilities but did not account for any nonlinear effects that might affect the resultant LCOs. The nonlinear phenomena that pertain to cochlear LCOs may include mutual suppression, the generation of harmonic and intermodulation distortion, and mode-locking as in musical wind instruments (Fletcher, 1999).

A. Aims and overview

This paper is concerned with the impact of cochlear nonlinearity upon the linear state space model's predictions of SOAEs. This work is an extension of the cochlear modeling findings presented in two previous papers. In the first, Elliott *et al.* (2007) introduced a state space framework for a classical model of the cochlea (Neely and Kim, 1986). In the second, Ku *et al.* (2008) applied the state space model to demonstrate how predictions of the coherent-reflection theory of SOAE generation (Shera and Zweig, 1993; Zweig and Shera, 1995) can be observed in a linear, non-scaling-symmetric model of the cochlea. The present paper investigates how linear instabilities interact and evolve in a nonlinear model of the cochlea to form LCOs similar to SOAEs.

A short review of the linear model is given in Sec. II. In order to better match the predicted variation of TW wavelength with position in the human cochlea, minor revisions to the model parameter values are presented. The implementation of a saturation nonlinearity in the model's CA is then described.

Section III presents the results from a variety of simulations. A set of linear stability tests is first presented. The results of a nonlinear time domain simulation involving a single linear instability, generated by introducing a step-change in feedback gain as a function of position, are then analyzed in detail and interpreted in terms of TWs. This is expanded upon by running a nonlinear simulation of a model with a small range of linear instabilities, generated by a spatially restricted region of inhomogeneities in the feedback gain. Finally, simulations of nonlinear cochleae with a large number of linear instabilities across the cochlea's frequency range are performed; these instabilities are generated by introducing inhomogeneities in the feedback gain all along the cochlea, as have been postulated to exist in the biological cochlea.

II. MODEL DESCRIPTION

The basis of the model used in this investigation is that of Neely and Kim (1986). Elliott *et al.* (2007) showed that it is possible to recast this frequency-domain system in state space. State space is a mathematical representation of a physical system that is commonly used in control engineering (Franklin *et al.*, 2005). One advantage of this formulation is that there are widely available, well-developed tools to study the behavior of such models. For example, it is straightforward to simulate state space models in the time domain using the ordinary differential equation solvers in MATLAB.

This section presents some minor revisions to the linear state space model and the necessary additions in order to account for CA nonlinearity in the time domain. Interested readers are directed to previous work for a more in-depth description of the model (Neely and Kim, 1986; Ku *et al.*, 2008; Ku, 2008) and the state space formulation (Franklin *et al.*, 2005; Elliott *et al.*, 2007).

A. The linear model

The active linear model seeks to represent a human cochlea's response to low-level stimuli where the CA is working at full strength. The Neely and Kim (1986) micromechanical model is based upon physical structures of the OC; it consists of a two-degree-of-freedom system with an active element in a feedback loop. This active element provides "negative damping" such that the cochlear TW due to a pure tone is enhanced basal of its peak. The original model described the cat cochlea; Ku *et al.* (2008) revised the parameters to account for the characteristics of a human cochlea. Minor modifications are provided here to reproduce the variation in TW wavelength as a function of position, as deduced from OAE data (Shera and Guinan, 2003). The parameters that have been updated from Ku *et al.* (2008) are

TABLE I. Model parameters for the revised quantities to represent a human cochlea.

Quantity	Revised formula (SI)	Units
$k_1(x)$	$1.65\times10^{9}e^{-279(x+0.00373)}$	N m^{-3}
$c_1(x)$	$9+9990e^{-153(x+0.00373)}$	N s m^{-3}
$m_1(x)$	4.5×10^{-3}	kg m^{-2}
$k_2(x)$	$1.05\times10^{7}e^{-307(x+0.00373)}$	N m^{-3}
$c_2(x)$	$30e^{-171(x+0.00373)}$	N s m^{-3}
$m_2(x)$	$7.20\times10^{-4}+2.87\times10^{-2}x$	kg m^{-2}
$k_3(x)$	$1.5\times10^{7}e^{-279(x+0.00373)}$	N m^{-3}
$c_3(x)$	$6.6e^{-59.3(x+0.00373)}$	N s m^{-3}
$k_4(x)$	$9.23\times10^{8}e^{-279(x+0.00373)}$	N m^{-3}
$c_4(x)$	$3300e^{-144(x+0.00373)}$	N s m^{-3}

shown in Table I. The micromechanical model and physical meaning of these quantities are described in Elliott *et al.* (2007).

The TW wavelength at its peak is approximately 0.5 mm at locations basal of 5 mm in the model; apical of this position, the TW wavelength increases linearly with position approaching 1.5 mm at the apex. The other TW characteristics such as enhancement and tonotopic tuning are broadly unchanged. A thorough review of the model and its responses to stimuli can be found in Ku (2008).

The macromechanical formulation of the state space model (Elliott *et al.*, 2007) is based upon the work of Neely (1981) and Neely and Kim (1986). A finite difference approximation is used to discretize the spatial derivatives in the wave equation and boundary conditions of the model. The local activity of the cochlear partition segments is related to the fluid mechanics by the one-dimensional wave equation in matrix form:

$$\mathbf{F}\mathbf{p}(t) - \ddot{\mathbf{w}}(t) = \mathbf{q}, \tag{1}$$

where $\mathbf{p}(t)$ and $\ddot{\mathbf{w}}(t)$ are the vectors of pressure differences and cochlear partition accelerations, $\mathbf{F}$ is the fluid-coupling matrix, and $\mathbf{q}(t)$ is the vector of source terms that serves as the input to the macromechanical model; in the baseline form of the model, $\mathbf{q}(t)$ is zero except at the stapes. The cochlear micromechanics of isolated partition segments are described by individual matrices. When Eq. (1) is substituted into an equation combining all the uncoupled elemental matrices, including the middle ear boundary at the base and the helicotrema at the apex, the dynamics of the fluid-coupled model can be described by the state space equations

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \tag{2}$$

and

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t), \tag{3}$$

where $\mathbf{A}$ is the system matrix that describes the coupled mechanics, $\mathbf{x}(t)$ is the vector of state variables which include BM and tectorial membrane (TM) velocities and displacements, $\mathbf{B}$ is the input matrix, $\mathbf{u}(t)$ is a vector of inputs equal to $\mathbf{F}^{-1}\mathbf{q}(t)$, $\mathbf{y}(t)$ is the output variable (BM velocity in this case), $\mathbf{C}$ is the output matrix, and $\mathbf{D}$ is an empty feed-through matrix. The details of this formulation are described by Elliott *et al.* (2007).

B. The nonlinear model

Including nonlinearity in a cochlear model can add complexity and greatly increase the computational intensity of a given simulation. However, there are many fundamental attributes of the cochlea that are not captured in a linear model. The primary source of saturating nonlinearity in the cochlea is the relative decrease in the OHC feedback force with increasing driving level; this effect is sometimes referred to as "self-suppression" in the literature (Kanis and de Boer, 1993).

To incorporate a saturation nonlinearity into the state space model, the feedback gain, γ, is made to depend on the instantaneous shear displacement between the BM and the TM, ξ_c. This can be expressed in state space as a time-varying version of the system matrix in Eq. (2), $\mathbf{A}(t)$, so that

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t), \tag{4}$$

where

$$\mathbf{A}(t) = \mathbf{A}_{\text{passive}} + \gamma[\xi_c(t)]\mathbf{A}_{\text{active}}. \tag{5}$$

The current model does not make any assumptions regarding the precise mechanisms of the electro-mechanical transduction in OHCs. The goal was to integrate the least complex saturation function into the model that also produces realistic results. A first order Boltzmann function is used to approximate the local saturation of the CA feedback force with level because its shape well approximates the input-output (pressure to intracellular voltage) characteristics of OHCs measured in isolation (e.g., Cody and Russell, 1987; Kros *et al.*, 1992). This function is applied to the displacement input of the feedback loop that determines the OHC force in the time domain:

$$f(u) = \delta\left(\frac{1}{1+\beta e^{-u/\eta}} - \frac{1}{1+\beta}\right), \tag{6}$$

where u is the input displacement, δ sets the saturation point, η affects the slope of the output, and β affects the asymmetry of the function. In order to set the slope of the function to unity for small input displacements, u, it is necessary to constrain

$$\eta = \frac{\delta\beta}{(1+\beta)^2}. \tag{7}$$

The free parameter, β, is set to 3. The function and its slope are illustrated in Fig. 1.

In order to determine $\gamma(t)$, the shear displacement waveform is passed through the Boltzmann function and scaled by the waveform itself:

$$\gamma(\xi_c(t)) = \left|\frac{f(\xi_c(t))}{\xi_c(t)}\right|. \tag{8}$$

The nonlinearity is thus both instantaneous and memory-less.

While the saturation point variable of the Boltzmann function, δ, is set to 1 in Fig. 1 for illustration purposes, it is varied as a function of position in this model of the nonlinear cochlea. Due to the decrease in BM stiffness with distance from the stapes, the cochlear partition will deflect more at the apex relative to the base when driven by a constant pressure.

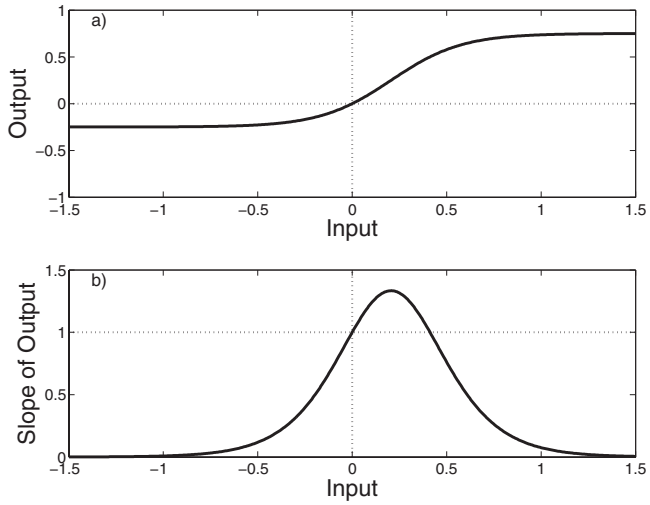


FIG. 1. Boltzmann function characteristics with saturation parameter $\delta=1$. (a) Output vs input. (b) Slope of output vs input.

Thus, $\delta(x)$ has a considerable impact upon the results of the simulation. In order to generate a sensible $\delta(x)$, the maximum displacement at a given location, calculated across frequencies in the coupled linear model, is used as a template for $\delta(x)$ at locations approximately 6 mm $\leq x \leq$ 27 mm; δ is fixed to constant values outside this range. This distribution is normalized to the maximum value and scaled by 10 nm; a saturation point that is approximately 1 nm at the base yields results that are similar to various experimental measurements made in animals (Robles and Ruggero, 2001). The final distribution of $\delta(x)$ is shown in Fig. 2.

III. Results

A. Linear stability tests

One of the advantages of the state space formulation is that it is possible to test the linear stability of a model. This is accomplished by calculating the eigenvalues of the system matrix, **A**, which correspond to the system poles (Franklin *et al.*, 2005). Ku *et al.* (2008) generated linear instabilities in the state space model by introducing inhomogeneities in the

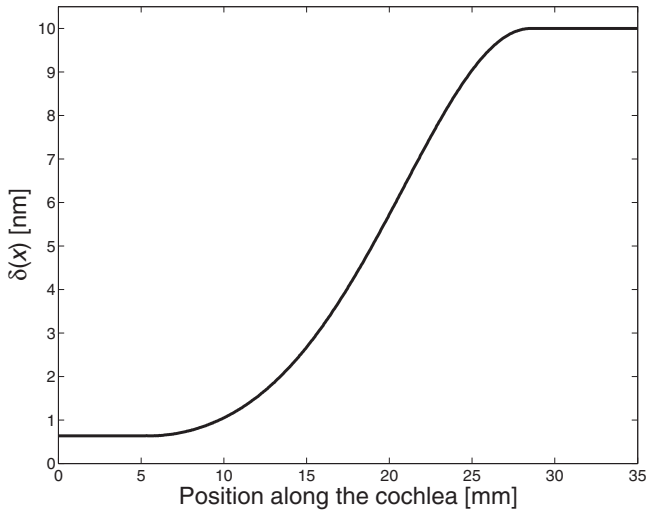


FIG. 2. Nonlinear saturation point of the micromechanical feedback loops as a function of position, $\delta(x)$.

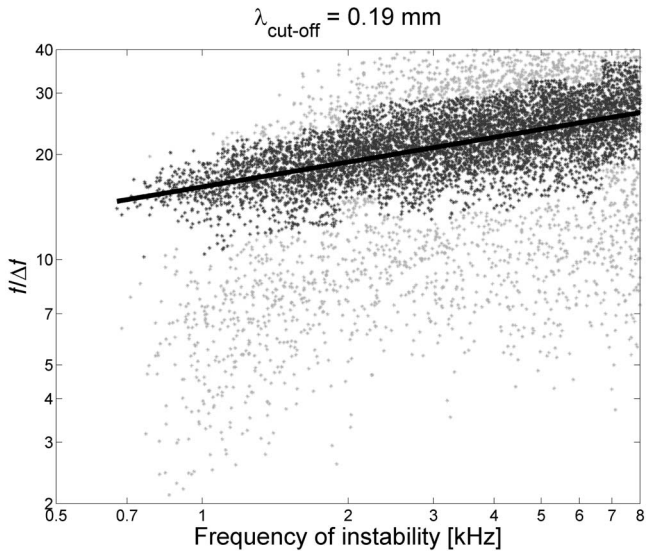


FIG. 3. Variation in the spacings between adjacent linear instabilities plotted against the geometric mean frequency of adjacent instabilities. The low-wavelength cut-off frequency of the random variations in $\gamma(x)$ is set to 0.19 mm. The darkened dots represent values of the spacings that fall within the ± 1 standard deviation of the mode within 15 log-spaced bands [see Fig. 3 of Shera (2003)]. A trend line is fitted to the modes of each band.

otherwise smooth distribution of $\gamma(x)$. The statistics of these instabilities satisfied a variety of predictions of the coherent reflection theory of SOAE generation. One aspect of human SOAE data that was not accounted for in the model of Ku *et al.* (2008) is the variation in SOAE spacings with frequency (Shera, 2003). This is addressed by the revised set of parameters presented in Table I.

Figure 3 represents the summarized stability tests of 200 cochlear models. Random "dense" distributions of $\gamma(x)$ were generated by filtering Gaussian white noise with fifth order Butterworth filters using a wide bandwidth of spatial frequencies (see Ku *et al.*, 2008; Lineton, 2001). In contrast to a "sparse" distribution of $\gamma(x)$, a dense distribution contains significant spatial wavenumber content at one-half the TW wavelength at its peak at all locations in the cochlea (Ku, 2008). The peak-to-peak variation in $\gamma(x)$ was held at a constant of 15%. The results of Fig. 3 show good qualitative agreement with statistics of measured SOAE spacings in humans (see Fig. 3 of Shera, 2003).

B. Nonlinear simulation: Step-change in $\gamma(x)$

In order to better understand the LCOs that develop in unstable nonlinear cochleae, the response of a model with a single linear instability is analyzed in detail. This instability is generated in the state space cochlea at 1.214 kHz by applying a step-distribution of feedback gain with $\gamma(x < 18.9 \text{ mm})=1$ and $\gamma(x \geq 18.9 \text{ mm})=0.85$. As shown in Fig. 4, a single pole exhibits a positive real part, $\sigma = 0.031 \text{ ms}^{-1}$, thus indicating that the system is unstable. It is also possible to create a model with a single unstable pole by applying a random $\gamma(x)$, though a step-distribution of $\gamma(x)$ has the added benefit that there is no ambiguity in the location of the reflection site.

The stability plot has been introduced previously as a convenient way of representing the poles of the system (El-

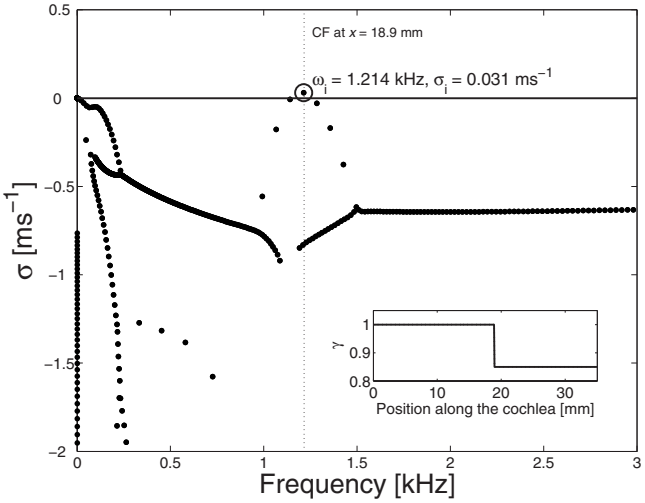


FIG. 4. Stability plot of a cochlear model with a step-change in $\gamma(x)$, shown in the inset. A vertical line marks characteristic frequency of a baseline cochlea at the location of the step-change in $\gamma(x)$. The unstable pole is circled and described by annotations.

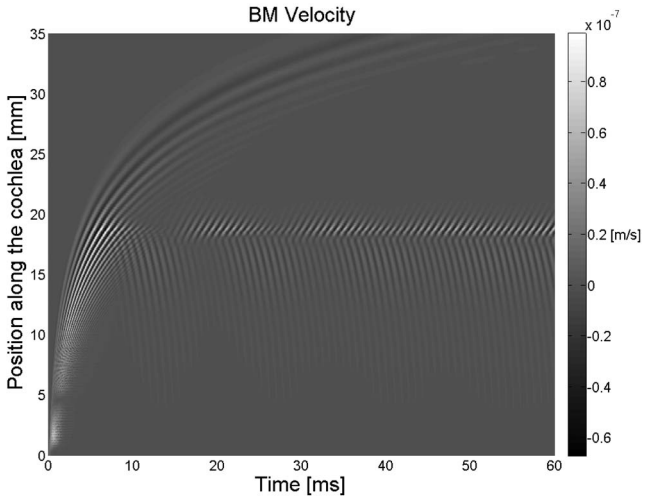


FIG. 5. Click response of the unstable nonlinear cochlear model described in Fig. 4. BM velocity is plotted against time and position along the cochlea.

liott *et al.*, 2007; Ku *et al.*, 2008). The imaginary part of a pole, plotted on the horizontal axis in kilohertz, represents its frequency; the real part of a pole, plotted on the vertical axis in inverse milliseconds [ms^{-1}], represents its rate of exponential growth or decay. Thus, a pole with a positive real part indicates that the system is unstable. Note that Fig. 4 only shows a small portion of the overall stability plot in order to emphasize the instability.

When the unstable model presented in Fig. 4 is excited in the nonlinear time domain by a click at the stapes, multiple TW reflections occur between the step-discontinuity in $\gamma(x)$ and the middle ear boundary condition to form a cochlear standing wave. This becomes a LCO over time due to the saturation of the CA. The cochlear response is plotted as a function of position and time in Fig. 5. Backward TWs are visible after the initial transient wavelet passes through the location of the discontinuity, $x=18.9$ mm. The spectrum of the pressure at the stapes also varies with time; this is of particular interest because it is understood that this pressure induces motion in the middle ear bones to generate OAEs in the ear canal. It is also possible to calculate the resulting pressure in the ear canal, though this is left to future work in order to simplify the interpretation of results.

The pressure spectrum at the base is plotted over four time windows in Figs. 6(a)–6(d). Superimposed on each frame is the stability plot over this frequency range. Blunt peaks are visible in the pressure spectrum of Fig. 6(a) at frequencies corresponding to the near-unstable poles as well as the single unstable frequency. The levels of these peaks also seem well-correlated with the relative magnitudes of the real parts of the poles, σ_i. This is consistent with the calculated linear transient response of a system with damped modes, which would include spectral components at each of the resonant frequencies of the system. The decay of these components is determined by the magnitude of the real parts of the corresponding poles, σ_i.

In this nonlinear system, distortion is visible at the second and third harmonics of the fundamentals. As later time frames are examined, only the response at the linearly unstable frequency and its harmonics persist. The frequency of the primary LCO in Fig. 6(d) is within 0.04% of the linear prediction, 1.214 kHz. The spectral resolution of the panels improves as longer time windows are analyzed in later time frames.

1. Interpretation in terms of TWs

The discrete Fourier transform (DFT) of the BM velocity is computed from $1000 \leq t \leq 3000$ ms at the frequency of the LCO, 1.214 kHz, for each location in the cochlea; the magnitude and phase are shown in Fig. 7. In the steady-state response of the model with a single linear instability due to a step-change in $\gamma(x)$, the peak in the TW phase as a function of position occurs basal to the peak in magnitude. This is

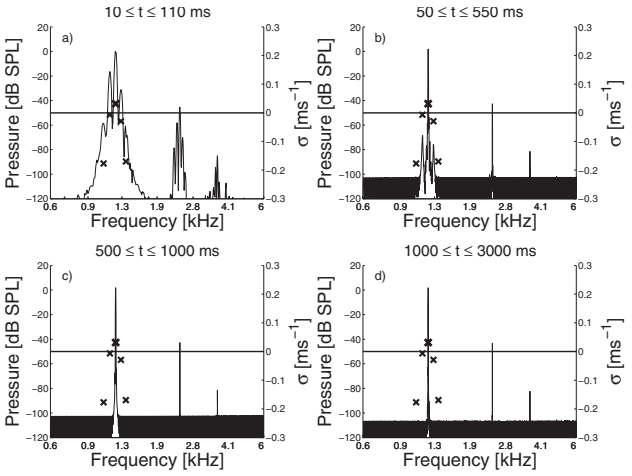


FIG. 6. Superimposed stability plots of the linear system (right axes) and the spectrum of the pressure at the base of a nonlinear, unstable cochlea (left axes) given four time windows: (a) $10 \leq t \leq 110$ ms; (b) $50 \leq t \leq 550$ ms; (c) 500 ms$\leq t \leq$1500 ms; (d) $1000 \leq t \leq 3000$ ms. The unstable pole is plotted with a dark **x**; stable poles are plotted with light x's.

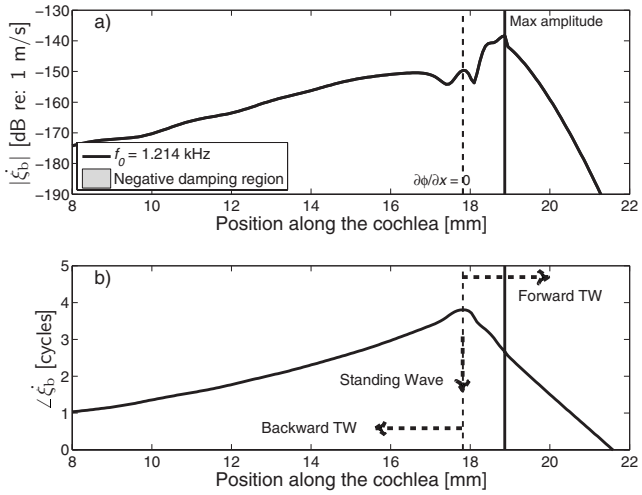


FIG. 7. Steady-state (a) magnitude and (b) phase of the BM velocity response as a function of position at the unstable frequency. A solid vertical line marks the location of the peak in magnitude at $x=18.9$ mm, and a dashed vertical line marks the location of the peak in the phase. The region of negative damping, determined by evaluating the real part of the BM impedance at a frequency of 1.214 kHz in a baseline active cochlea, is shaded.

sometimes interpreted to mean that the TW is reflected basal to the location of the impedance discontinuity, resulting in a roundtrip delay that is less than twice the forward delay. The actual physical meaning of the data can be clarified by examining this result in closer detail.

The dominant direction of wave propagation in this cochlear model is indicated by the sign of the slope of the phase response as a function of position, $\partial\phi/\partial x$. For instance, a positive phase slope indicates a predominantly backward TW, a zero phase slope indicates a standing wave, and a negative phase slope indicates a predominantly forward TW. It is not a coincidence that the position where the phase slope is zero is located within the negative damping region at this frequency, which extends approximately 2 mm basal and 1 mm apical of the $\partial\phi/\partial x=0$ place. The imposed step-discontinuity in $\gamma(x)$ coincides with both the apical end point of the negative damping region, shown in Fig. 7 as a shaded area, and the location of maximum TW magnitude.

When the initial forward traveling wavelet passes through the negative damping region, it is amplified at the frequency of the instability. This response peaks at the location of the step-discontinuity, which causes a wavelet to be reflected back toward the base. As this reflected wavelet passes back through the negative damping region, it is again amplified. The amplification of backward TWs in a one-dimensional model has been previously demonstrated by Talmadge *et al.* (1998). The peak in the phase response represents the only position along the BM where the amplitudes of forward and backward TWs are equal; as a result, the TW is "standing" at the $\partial\phi/\partial x=0$ location. The nearby ripples in the magnitude are likely due to constructive and destructive interference of forward and backward TWs; this is only visible where these TW components are approximately equal in amplitude.

Perhaps contrary to one's intuition, the roundtrip TW delay is still twice the forward delay. However, this is no longer apparent from the phase plot as the backward TWs are hidden "under" the dominant forward TWs in the overlap region basal to the characteristic place. Similar plots are generated when forward TWs are reflected by spatially random inhomogeneities in linear wave models (Neely and Allen, 2009).

2. Linear analysis of the steady-state nonlinear response

Given both steady-state pressure and velocity, it is possible to reconstruct the effective linear feedback gain as a function of position in this nonlinear simulation, $\gamma_{\text{eff}}(x)$; this line of reasoning follows from de Boer's (1997) EQ-NL theorem. Solving the BM impedance equation (Eq. 12 in Neely and Kim, 1986) for the effective feedback gain yields

$$\gamma_{\text{eff}}(x)\big|_{f=1.214\ \text{kHz}} = \left.\frac{Z_1Z_2 + Z_1Z_3 + Z_2Z_3 - (Z_1+Z_2)[P(x)/\dot{\xi}_b(x)]}{Z_2Z_4}\right|_{f=1.214\ \text{kHz}}, \tag{9}$$

where $P(x)$ is the pressure difference across the cochlear partition, $\dot{\xi}_b(x)$ is the BM velocity, and Z_i are the frequency- and position-dependent impedances of the micromechanical model (Neely and Kim, 1986). The distribution of $\gamma_{\text{eff}}(x)$ that is calculated from the above nonlinear simulation at steady-state is presented in Fig. 8, in addition to the original imposed $\gamma(x)$. In the region just basal to the TW peak, where negative damping takes place, $\gamma_{\text{eff}}(x)$ is less than the imposed gain. The effective gain calculation breaks down beyond $x > \sim 23$ mm, where the TW becomes evanescent.

It is informative to now apply $\gamma_{\text{eff}}(x)$ to a linear state space cochlea to study its stability, as shown in Fig. 9. γ_{eff} is set=0.85 for $x>23$ mm in this linear stability test. The pole located at 1.214 kHz is no longer unstable, but its real part is very nearly zero. The effective linear system will ring at 1.214 kHz with a velocity distribution that is almost identical to that of Fig. 7, but its oscillations will neither grow nor decay significantly with time; this is consistent with the observed LCO behavior in the nonlinear simulation. From this, it is possible to deduce that the amount of work done by the CA at this frequency is equal to the losses in the cochlear model, as predicted by the coherent-reflection theory (Shera, 2003).

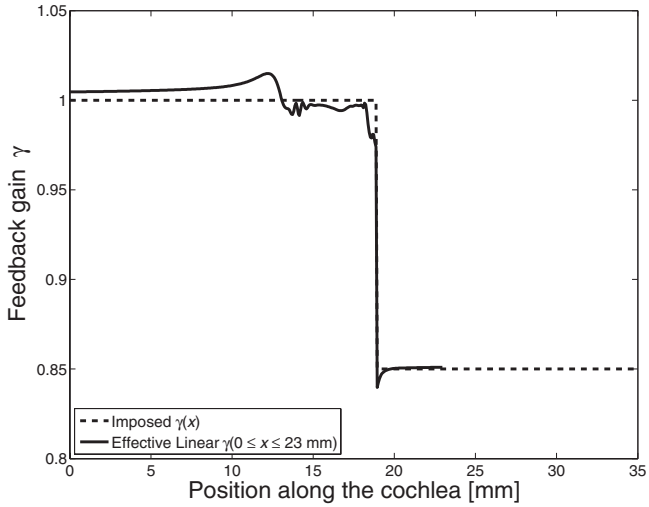


FIG. 8. The effective linear feedback gain of the nonlinear simulation at steady-state as a function of position (—) and the original imposed feedback gain (– –).

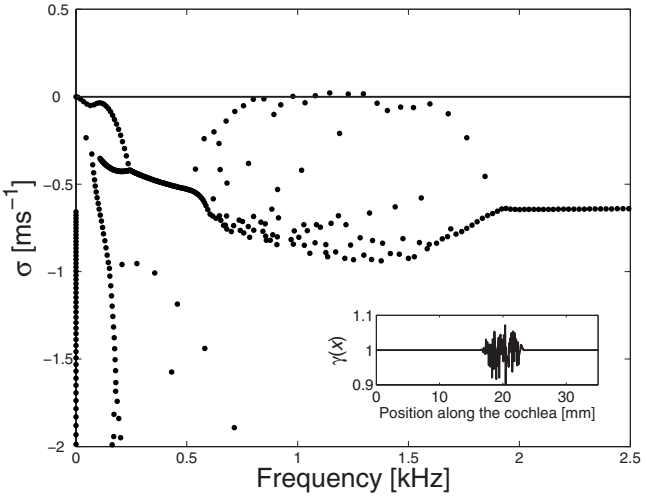


FIG. 10. Stability of a linear cochlear model given a windowed-perturbed gain distribution, shown in the inset. Five unstable poles result.

C. Nonlinear simulations: Random changes in $\gamma(x)$

When multiple linear instabilities are predicted, it is possible that nonlinear interactions between the resultant LCOs will suppress one another or generate intermodulation distortion. This subsection presents simulations of cochleae with dense inhomogeneities that express a broad band of spatial frequencies, as in Ku *et al.* (2008). In the first instance, a linearly unstable system with a spatially limited region of inhomogeneity is generated to study the interactions of a small number of instabilities. Subsequently, systems with inhomogeneities distributed along the lengths of the cochlear models are examined. In each case, the spectral content of LCOs in nonlinear time domain simulations is compared against the frequencies of the linear instabilities.

1. Isolated region of inhomogeneity

In order to restrict the inhomogeneous region in space, a 3.5 mm long Hann window is extended by inserting 3.5 mm ones at its midpoint; zeros are padded outside this range, and the function is centered at $x=19$ mm. By applying this extended Hann window to a dense distribution of random inhomogeneities in $\gamma(x)$, five linear instabilities are generated at $f=[0.979, 1.080, 1.145, 1.229, 1.296]$ kHz, as shown in Fig. 10. As before, the nonlinear pressure spectrum at the base of the cochlear model is calculated across four time windows and compared with the linear stability plot in Fig. 11.

The evolution of the nonlinear response shown in Fig. 11 is analogous to that of Fig. 6. In the earliest time frame [Fig. 11(a)], there are peaks in the pressure spectrum at all of the resonant modes generated by the inhomogeneities—both stable and unstable. However, as time progresses, the peaks at the near-unstable frequencies die away. In contrast to the case of the isolated instability, not all of the linearly unstable frequencies are represented at steady-state, as seen in Fig. 11(d). Peaks are also detected at frequencies that correspond

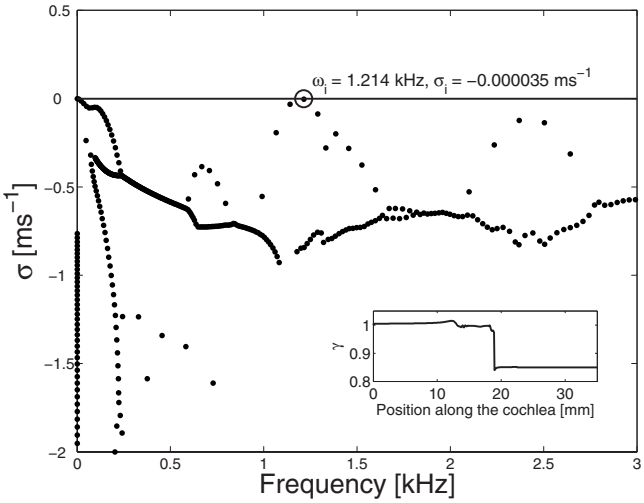


FIG. 9. Linear stability plot as determined by applying the effective linear gain as a function of position shown in Fig. 8. Note that the applied gain has been set to 0.85 for $x>23$ mm.

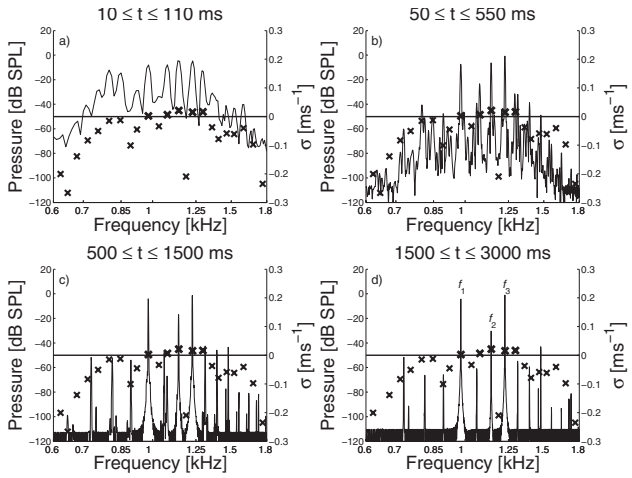


FIG. 11. Simultaneous plots of the linear system stability (right axes) and the spectrum of the pressure at the base of the nonlinear cochlear model (left axes) given four time windows: (a) $10\leq t\leq 110$ ms; (b) $50\leq t\leq 550$ ms; (c) $500\leq t\leq 1500$ ms; (d) $1500\leq t\leq 3000$ ms. Unstable poles are plot with dark **x**'s; stable poles are plot with light x's. Three linearly unstable frequencies, f_1, f_2, and f_3, are indicated.

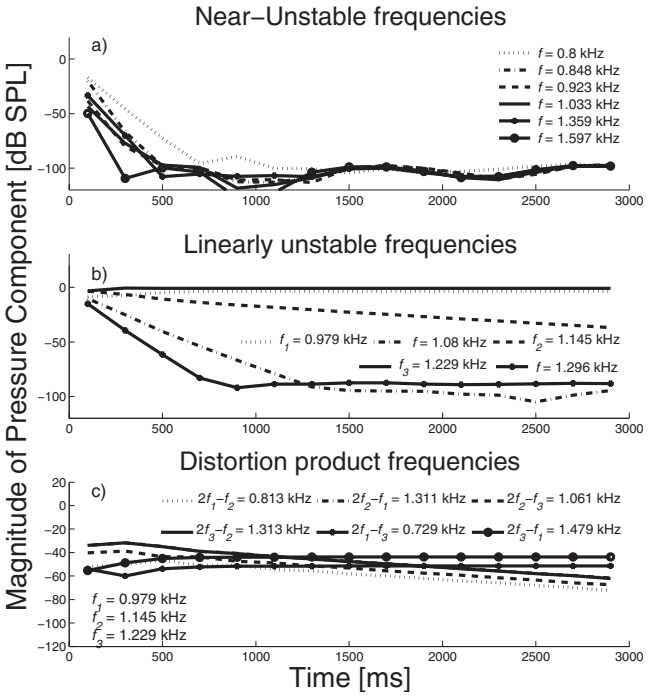


FIG. 12. Variation in the magnitude of basal pressure frequency components in an unstable cochlea as a function of time. Near-unstable and linearly unstable frequencies are shown in (a) and (b), respectively, while distortion product frequencies are shown in panel (c). Every curve consists of 15 data points, where each value represents the DFT of 200 ms of data with no overlap between adjacent windows.

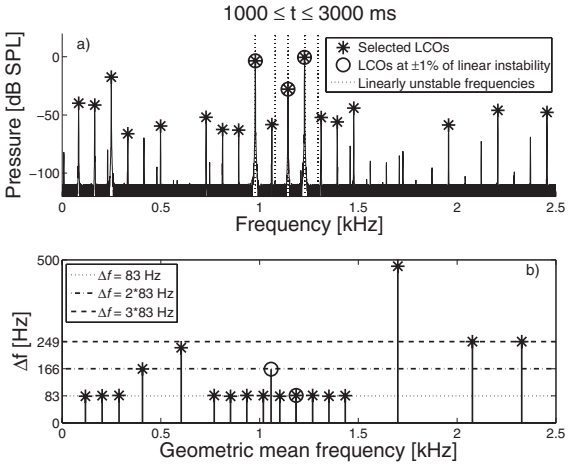


FIG. 13. The steady-state basal pressure spectrum is displayed across a limited frequency range in panel (a). A star (∗) marks selected LCOs. Selected LCOs that also correspond to linearly unstable frequencies and spacings are denoted by a circle (○). The horizontal axis of panel (b) represents the geometric mean of the two adjacent limit cycle frequencies; the vertical axis represents the separation between LCOs in Hz. Dotted, dot-dashed, and dashed horizontal lines are drawn at $\Delta f = [83, 2\times83, 3\times83]$ Hz, respectively.

to intermodulation distortion between the persistent linearly unstable LCOs. These trends are summarized and clarified by Fig. 12, which plots the variation in the magnitudes of three sets of frequency components with time.

The magnitudes of the basal pressure response at the near-unstable frequencies in Fig. 12(a) all drop away into the simulation's noise floor within approximately 500 ms of the initial stimulus. This rate of exponential decay is roughly −150 dB/s. The magnitudes of three of the five linearly unstable frequencies, located at 1.08, 1.145, and 1.296 kHz, also decay away with time, as shown in Fig. 12(b). This is believed to be due to mutual suppression of the LCOs. However, the decay rate of these components is markedly slower than that of the near-unstable frequencies. For instance, the $f=1.296$ kHz and $f=1.080$ kHz magnitudes initially decay away at approximately −100 and −70 dB/s, while the $f=1.145$ kHz component diminishes much more slowly at roughly −10 dB/s. Only two of the linearly unstable frequencies, located at $f=1.229$ kHz and $f=0.979$ kHz, evolve into stable LCOs; the magnitudes of these persistent components stabilize at −0.8 and −3.5 dB SPL, respectively.

The magnitudes of a number of commonly observed distortion products which result from three assumed primaries are given in Fig. 12(c). In addition to the most commonly studied distortion product OAE (DPOAE), the "lower" cubic distortion product ($2f_l - f_h$), another nearby DPOAE ($2f_h - f_l$) is also examined. General notations of f_l and f_h, corresponding to the frequencies of the lower tone and the higher tone, are adopted above to avoid confusion with the notation for the selected primaries. The primaries chosen are the three linearly unstable frequencies that persist in amplitude: $f_1 = 0.979$ kHz, $f_2 = 1.145$ kHz, and $f_3 = 1.229$ kHz. The magnitudes of the distortion products at $2f_1 - f_3$ and $2f_3 - f_1$ mirror the persistence of the two primaries at f_1 and f_3, just as the magnitudes of the other four distortion products show slow decay, in a manner similar to f_2. Note, however, that decay rates of these distortion products are somewhat less steep than that of f_2; this is perhaps because the amplitudes of the other primaries are stable.

The time-varying SOAE phase analysis described by van Dijk and Wit (1998) was performed upon the hypothesized distortion product LCOs (DPLCOs) shown in Fig. 12(c). It was determined with a high degree of confidence that these LCOs were all phase-locked to their assumed primaries with the exception of $2f_2 - f_3 = 1.061$ kHz and $2f_2 - f_1 = 1.311$ kHz. The latter case can be explained because another nearby DPLCO, $2f_3 - f_2 = 1.313$ kHz, appears to have suppressed the 1.311 kHz signal; the DPLCO at 1.313 kHz is indeed phase-locked to $2\varphi_{f_3} - \varphi_{f_2}$. It is possible that the LCO at 1.061 kHz is not locked to $2\varphi_{f_2} - \varphi_{f_3}$ given its proximity to a linear instability at 1.080 kHz. One might also observe that $f_1 \approx 3f_2 - 2f_3$ within 0.1%. However, the phase analysis shows that the LCOs at f_1, f_2, and f_3 are in fact independent of one another as $\varphi_{f_1} - (3\varphi_{f_2} - 2\varphi_{f_3})$ varies continuously with time; one explanation is that the distortion generated at roughly this frequency is being entrained by the underlying linear instability at f_1.

One of the salient features of mammalian SOAEs is the distribution of spacings between unstable frequencies. The steady-state basal pressure spectrum of this system is plotted in Fig. 13(a). To examine the log-normalized spacings between adjacent limit cycles, an arbitrary threshold was set at −65 dB below the strongest instability to choose frequencies for analysis. The resultant Δf spacings of the selected limit cycles are shown in Fig. 13(b). The direct spacings between

selected LCOs are plotted instead of the log-normalized spacings to emphasize the harmonic nature of the spacings.

2. Inhomogeneities throughout the cochlear model

When random dense inhomogeneities are introduced along the entire length of the cochlear model, linear instabilities can be generated across its whole frequency range. In this subsection, 3 s long nonlinear time domain simulations are performed on 20 linearly unstable cochlear models. The peak-to-peak variations in $\gamma(x)$ range logarithmically from 2% to 20%. Between 100 and 160 h are required to compute a single 3 s long nonlinear time domain simulation, where smaller peak-to-peak variations in $\gamma(x)$ required less time. Even at the highest applied values of feedback gain, all of the isolated micromechanical models remain stable; the system only becomes unstable when all of the cochlear elements are coupled together by the fluid.

Figure 14 I and II show the results for two of these cochlear models. The peak-to-peak variations in $\gamma(x)$ are 3.67% and 10.9%, respectively. Note that the apparent steady-state "noise floor" rises as the peak-to-peak variation in $\gamma(x)$ increases. However, the error computational tolerances are held constant for all 20 simulations, indicating that the apparent noise is due to cochlear activity. The (a) panels of Fig. 14 I and II show the steady-state basal pressure spectrum in detail over a small frequency range. A number of LCOs are selected to be analyzed in terms of their adjacent log-normalized spacings.

In Figs. 14 Ib and IIb, the spacings between adjacent selected LCOs are plotted as a function of the geometric mean of a given pair of LCOs; the same notation as in the (a) panels is preserved, but diamonds (◇) are also included to represent the spacings between adjacent linear instabilities. Some spacings between LCOs are marked by all symbols—a circle, a diamond, and an asterisk—thus indicating that two adjacent LCOs both correspond to linear instabilities. Though there are a number of such near-overlaps, this is more often the exception than the rule.

The best process of determining what qualifies as a SOAE given an experimental measurement has been previously debated within the literature (e.g., Talmadge and Tubis, 1993; Zhang and Penner, 1998; Pasanen and McFadden, 2000). These methods seek to isolate and identify SOAEs from physiological background noise. In the case of the present simulations, it can be argued that every steady-state oscillation is potentially a SOAE: The only "true" noise in the system is due to simulation error, which is well below the magnitude of the LCOs. Thus, the challenge here is not to identify what signals originate in the cochlear model, but rather which LCOs might be detected and labeled as SOAEs in the ear canal.

Peaks in the spectrum are identified by comparing a given frequency magnitude to the magnitude at the adjacent lower frequency. If the magnitude increases by a certain threshold, the frequency at which the highest local magnitude occurs is selected as a possible SOAE. The chosen threshold decreases linearly from 70 to 35 dB as a function of logarithmically increasing peak-to-peak variation in $\gamma(x)$.

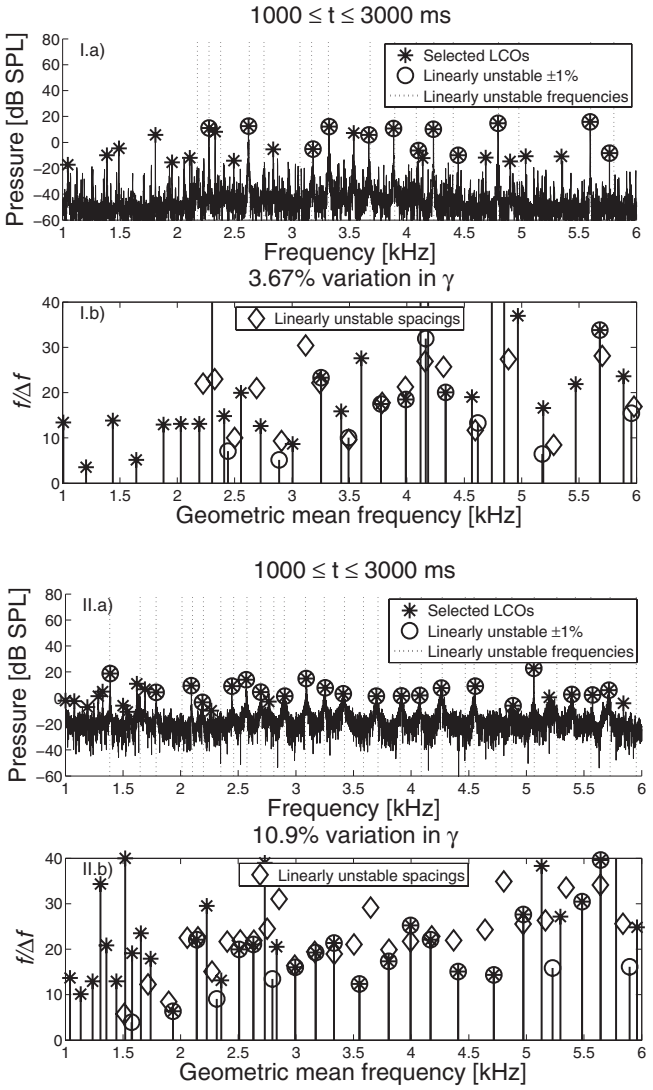


FIG. 14. Steady-state basal pressure spectrum (a) with further annotations: selected limit cycles are marked with a star (*), whereas those that fall within ±1% of a linear instability (dotted vertical line) are marked with a circle (○). Panel (b) shows the spacings between selected LCOs with the same marking conventions as (a). Diamonds (◇) mark the spacings between adjacent linear instabilities. Peak-to-peak variations in $\gamma(x)$ are 3.67% in the first set (I) and 10.9% in the second set (II).

If this threshold is set too low, an unrealistic number of sharp peaks are detected in models with small peak-to-peak variations in $\gamma(x)$; if this threshold is set too high, almost no peaks are detected in models with large peak-to-peak variations in $\gamma(x)$. This is due to the changing level of the apparent noise floor with peak-to-peak variations in $\gamma(x)$. It is not claimed that this method is optimal or ideal, but it represents a first attempt to compare the spacings between nonlinear LCOs to those of the linear instabilities. Further consideration of this topic is given in the discussion.

The spacing data from all 20 simulations are plotted simultaneously in Fig. 15. Panel (a) shows the spacings of all the linear instabilities from these models, whereas panel (b) shows all the spacings from selected nonlinear LCOs from these simulations. While the spacings between linear instabilities show a relatively tight banding, the spacings between nonlinear LCO frequencies are much more sparsely spread

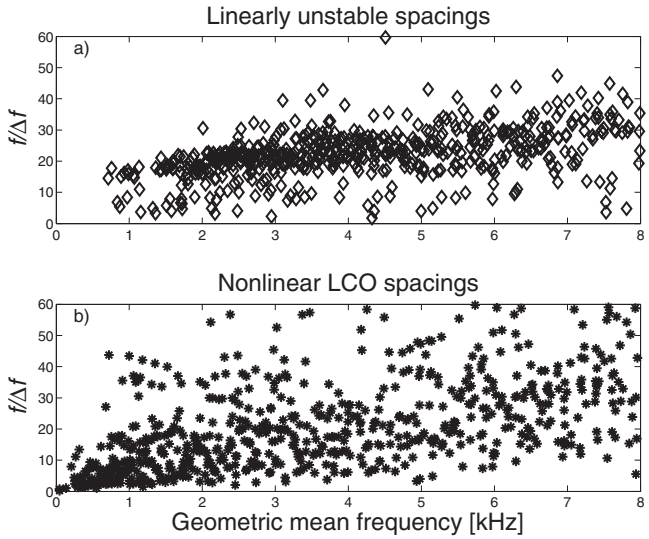


FIG. 15. (a) Log-normalized spacings between adjacent linear instabilities and (b) spacings between selected adjacent, steady-state nonlinear LCOs in 20 linearly unstable cochlear models.

apart. Compared to Fig. 13(b), there is far less uniformity in Fig. 15(b); this suggests that the interactions of LCOs in models with inhomogeneities at all x are less local and therefore more complex. There is also a large cluster of nonlinear LCO spacings at low mean frequencies and $f/\Delta f$ spacings in Fig. 15(b) that do not correspond to the frequencies of linear instabilities. These LCOs appear to be due to difference tones generated by higher-frequency "primary" oscillations. The $f_3-f_2=0.083$ kHz LCO shown in Fig. 13 is an example of such a difference tone.

IV. DISCUSSION AND CONCLUSIONS

The state space model of the cochlea has been applied to study the nonlinear activity of linear instabilities. Isolated instabilities clearly exhibit the features of Zweig and Shera (1995) coherent-reflection theory of SOAE generation. However, the oscillations due to linear instabilities can interact to suppress one another or generate intermodulation and harmonic distortion.

The nonlinear simulations presented here have illustrated how SOAEs may evolve in the biological cochlea as LCOs. In the simplest system, a step-distribution of $\gamma(x)$ was applied to a cochlear model that generated a single linear instability. The resultant nonlinear steady-state response only oscillated at the frequency of the linear instability and its harmonics. The magnitudes of the first and second harmonics were approximately 40 and 80 dB down from that of the fundamental tone. The form of the saturation nonlinearity applied to the feedback loop strongly affects these values. For instance, a symmetrical function such as the hyperbolic tangent will not generate any even-order distortion.

When multiple linear instabilities are present in a nonlinear simulation, the principles of the coherent-reflection theory are applicable but with some important additional considerations. As seen in Fig. 12(b), oscillations at linearly unstable frequencies can be suppressed over time. The rate of this suppression can vary from instability to instability. However, the nonlinear suppression of linear instabilities takes place over much longer time scales than the exponential decay of the linearly near-unstable modes.

Sharply tuned but diminishing tones, similar to the behavior of the oscillations at near-unstable frequencies, have been detected in the course of measuring the spectra of human SSOAEs; these are referred to as "decaying components" (Sisto *et al.*, 2001; Jedrzejczak *et al.*, 2008). These authors reported decay rates on the order of −500 dB/s. Much shorter time epochs (around 80 ms) are commonly used in experimental measurements of SSOAE to maintain high signal-to-noise ratios when averaging over many recordings, relative to the 3000 ms simulations. When the initial decay rates of the near-unstable modes shown in Fig. 12(a) are examined in detail over the first 100 ms following the click stimulus (data not shown), decay rates ranging from approximately −200 to −500 dB/s are found. This suggests that the measured SSOAE decaying components are likely due to stable resonances caused by multiple reflections between the base and apical inhomogeneities.

Assuming that the real cochlea behaves as the model, it should be possible to record the decay of oscillations due to click-stimulated linear instabilities that are suppressed over a matter of seconds, given ears with multiple SOAEs. Unfortunately, the extended recording periods required make this long SSOAE measurement less attractive in practice. One would also expect the existence of SOAEs that are generated by intermodulation distortion between two primary SOAEs, as first reported by Burns *et al.* (1984). Applying the methodology outlined by van Dijk and Wit (1998) showed that most of the presumed third order DPLCOs in Fig. 12(c) were indeed locked to their linearly unstable primaries (data not shown). van Dijk and Wit (1998) argued that a DPLCO may entrain an independent LCO. This phenomenon was not observed in the linearly unstable modes corresponding to f_1, f_2, and f_3 in Fig. 12(a); it may be that the fifth order intermodulation distortion generated by f_2 and f_3 was not strong enough to entrain the linearly unstable mode at f_1.

One question of interest is whether the spacings between adjacent linear instabilities are commensurate with the spacings between adjacent nonlinear LCOs. Unfortunately, at this time, there is no direct method of ascertaining which linear instabilities will evolve into persistent LCOs except by nonlinear simulation. The primary difficulty associated with interpreting the spacings between LCOs is how to decide whether a given peak in the basal pressure spectrum would be labeled as a SOAE. The somewhat *ad hoc* method applied in this paper represents an initial attempt to interpret the results.

Perhaps a more robust method for selecting nonlinear LCOs as potential SOAEs would involve transforming the pressure at the base to the pressure in the ear canal. The LCOs that protrude above the physiologically measured noise floor would then be designated SOAEs. This may reduce the number of selected LCOs at very low frequencies, for instance, due to the band-pass nature of the reverse middle ear transfer function (Puria, 2003). However, the best-attainable noise floor of experimental measurements of SOAEs is a function that varies with quantities such as fre-

quency and processing technique. In addition, there is no guarantee that model parameters such as the saturating point as a function of position, $\delta(x)$, accurately reflect the biophysics of the cochlea. As such, the complexity of the post-processing techniques must be balanced against the current uncertainties in the modeling of the system. Further refinements in this area are left to future work, though some general conclusions can be made from these simulations.

Most LCOs presented in Fig. 14 fall within 1% of the frequency of pre-existing linear instabilities. However, there are gaps where linear instabilities are suppressed. There is also a significant subset of LCOs that exist at frequencies not predicted by linear analysis. This is most visible in Fig. 14 Ia, where there are several frequency ranges that are free of linear instability. For instance, the LCOs at $f<2$ kHz, $f \approx 3.5$ kHz, and $4.5<f<5.5$ kHz all lack linear counterparts. The spacings between these LCOs, as shown in Fig. 14 Ib, are all similar to those of the nearby linearly predicted spacings.

Consider that the nearest distortion product frequencies of two hypothetical primary LCOs at 1.0 and 1.1 kHz would be located at 0.9 and 1.2 kHz. This would result in Δf spacings that fall at regular intervals. Even if there is a region free of LCOs that correspond to linear instabilities, perhaps due to nonlinear suppression or a smoother local variation of cochlear partition impedance, it is likely that a distortion product will be generated nearby due to the next higher two (or previous lower two) instabilities in frequency. Although linear reflection is no longer the mechanism giving rise to all LCOs, the local spacings predicted by the coherent-reflection theory would still be expressed. This phenomenon nevertheless requires that a dense distribution of inhomogeneities be present in the first place to fix the regular underlying spacings between linear instabilities, as dictated by the TW wavelength at its peak and the frequency-to-place map. Thus, the model predicts that human SOAEs are amplitude-stabilized cochlear standing waves and their intermodulation distortion products.

When spectra surrounding individual LCOs are examined, the sharpness of the peak varies from one simulation to the next. This is related to the level of the apparent noise floor, which varies proportionally with the peak-to-peak variations in $\gamma(x)$. Presumably, the increase in cochlear noise with the magnitude of the inhomogeneities is a result of an increasing number of linear reflections generating LCOs, each generating its own distortion. This cochlear noise appears to widen the spectral widths of LCOs in a manner that is analogous to the effect of random noise upon the spectral widths of van der Pol oscillators (e.g., Bialek and Wit, 1984; Long *et al.*, 1991, 1996; van Dijk and Manley, 2009).

Nonlinear simulations are capable of revealing intricacies of the cochlea not predicted by linear analysis. The findings presented in this paper reinforce the need to take the saturation of the CA into account when dealing with inherently nonlinear phenomena such as SOAEs; this is facilitated by the use of the state space formulation. In this context, nonlinear simulations are necessary as there is currently no method to predict which linear instabilities will develop into persistent LCOs and thus potential SOAEs. It is hoped that more researchers will adopt nonlinear models as there are still many cochlear phenomena to explore.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Chris Shera for some stimulating conversations regarding cochlear mechanics. We are also grateful for the insightful suggestions provided by two anonymous reviewers.

Allen, J. B., Shaw, G., and Kimberley, B. P. (**1995**). "Characterization of the nonlinear ear canal impedance at low sound levels," Assoc. Res. Otolaryngol. Abstr. **18**, 797.

Ashmore, J. F. (**1987**). "A fast motile response in guinea-pig outer hair cells: The cellular basis of the cochlear amplifier," J. Physiol. (London) **388**, 323–347.

Bialek, W., and Wit, H. (**1984**). "Quantum limits to oscillator stability: Theory and experiments on acoustic emissions from the human ear," Phys. Lett. **104**, 173–178.

Braun, M. (**1997**). "Frequency spacing of multiple spontaneous otoacoustic emissions shows relation to critical bands: A large-scale cumulative study," Hear. Res. **114**, 197–203.

Burns, E. M., Strickland, E. A., Tubis, A., and Jones, K. (**1984**). "Interactions among spontaneous otoacoustic emissions. I. Distortion products and linked emissions," Hear. Res. **16**, 271–278.

Cody, A. R., and Russell, I. J. (**1987**). "The responses of hair cells in the basal turn of the guinea pig cochlea to tones," J. Physiol. (London) **383**, 551–569.

Dallmayr, C. (**1985**). "Spontane oto-akusticsche Emissionen: Statistik und Reaktion auf akustische Störtöne (Spontaneous oto-acoustic emission: Statistics and the reaction to acoustic suppression tones)," Acustica **59**, 67–75.

Dallmayr, C. (**1986**). "Stationäre und dynamische Eigenschaften spontaner und simultan evozierter oto-akustischer Emissionen (Static and dynamic properties of spontaneous and simultaneously evoked oto-acoustic emissions)," Ph.D. dissertation, Technische Universität, Munich.

de Boer, E. (**1997**). "Connecting frequency selectivity and nonlinearity for models of the cochlea," Aud. Neurosci. **3**, 377–388

Elliott, S. J., Ku, E. M., and Lineton, B. (**2007**). "A state space model for cochlear mechanics," J. Acoust. Soc. Am. **122**, 2759–2771.

Fletcher, N. H. (**1999**). "The nonlinear physics of musical instruments," Rep. Prog. Phys. **62**, 723–764.

Franklin, G. F., Powell, J. D., and Emani-Naeini, A. (**2005**). *Feedback Control of Dynamics Systems*, 5th ed. (Addison Wesley, Reading, MA).

Gelfand, S. A. (**1998**). *Hearing: An Introduction to Psychological and Physiological Acoustics* (Marcel Dekker, New York).

Gobsch, H., and Tietze, G. (**1993**). "Interrelation of spontaneous and evoked otoacoustic emissions," Hear. Res. **69**, 176–181.

Gold, T. (**1948**). "Hearing. II. The physical basis of the action of the cochlea," Proc. R. Soc. London, Ser. B **135**, 492–498.

Hall, J. W. (**2000**). *Handbook of Otoacoustic Emissions* (Singular Publishing Group, San Diego).

Jedrzejczak, W. W., Blinowska, K. J., Kochanek, K., and Skarzynski, H. (**2008**). "Synchronized spontaneous otoacoustic emissions analyzed in a time-frequency domain," J. Acoust. Soc. Am. **124**, 3720–3729.

Kanis, L. J., and de Boer, E. (**1993**). "Self-suppression in a locally active nonlinear model of the cochlea: A quasi-linear approach," J. Acoust. Soc. Am. **94**, 3199–3206.

Kemp, D. T. (**1979**). "Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea," Arch. Oto-Rhino-Laryngol. **224**, 37–45.

Kros, C. J., Rüsch, A., and Richardson, G. P. (**1992**). "Mechanoelectrical transducer currents in hair cells of the cultured mouse cochlea," Proc. R. Soc. London, Ser. B **249**, 185–193.

Ku, E. M. (**2008**). "Modelling the human cochlea," Ph.D. thesis, University of Southampton, Southampton, UK.

Ku, E. M., Elliott, S. J., and Lineton, B. (**2008**). "Statistics of instabilities in a state space model of the human cochlea," J. Acoust. Soc. Am. **124**, 1068–1079.

Lineton, B. (**2001**). "Testing a model of the stimulus frequency otoacoustic

emissions in humans," Ph.D. thesis, University of Southampton, Southampton, UK.
Long, G. R., Tubis, A., and Jones, K. (**1991**). "Modelling synchronization and suppression of spontaneous otoacoustic emissions using van der pol oscillators: Effects of aspirin administration," J. Acoust. Soc. Am. **89**, 1201–1212.
Long, G. R., van Dijk, P., and Wit, H. P. (**1996**). "Temperature dependence of spontaneous otoacoustic emissions in the edible frog (Rana esculenta)," Hear. Res. **98**, 22–28.
Murphy, W. J., Tubis, A., Talmadge, C. L., and Long, G. R. (**1995**). "Relaxation dynamics of spontaneous otoacoustic emissions perturbed by external tones. II. Suppression of interacting emissions," J. Acoust. Soc. Am. **97**, 3711–3720.
Neely, S. T. (**1981**). "Finite difference solution of a two-dimensional mathematical model of the cochlea," J. Acoust. Soc. Am. **69**, 1386–1393.
Neely, S. T., and Allen, J. B. (**2009**). In *Concepts and Challenges in the Biophysics of Hearing*, edited by N. P. Cooper and D. T. Kemp (World Scientific, Singapore), pp. 62–67.
Neely, S. T., and Kim, D. O. (**1986**). "A model for active elements in cochlear biomechanics," J. Acoust. Soc. Am. **79**, 1472–1480.
Pasanen, E. G., and McFadden, D. (**2000**). "An automated procedure for identifying spontaneous otoacoustic emissions," J. Acoust. Soc. Am. **108**, 1105–1116.
Penner, M. J., and Zhang, T. (**1997**). "Prevalence of spontaneous otoacoustic emissions in adults revisited," Hear. Res. **103**, 28–34.
Pickles, J. O. (**2003**). *An Introduction to the Physiology of Hearing*, 2nd ed. (Academic, London).
Probst, R., Coots, A. C., Martin, G. K., and Lonsbury-Martin, B. L. (**1986**). "Spontaneous, click-, and toneburst-evoked otoacoustic emissions from normal ears," Hear. Res. **21**, 261–275.
Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2066.
Puria, S. (**2003**). "Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions," J. Acoust. Soc. Am. **113**, 2773–2789.
Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.
Shera, C. A. (**2003**). "Mammalian spontaneous otoacoustic emissions are amplitude-stabilized cochlear standing waves," J. Acoust. Soc. Am. **114**, 244–262.
Shera, C. A., and Guinan, J. J. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.
Shera, C. A., and Guinan, J. J. (**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., 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.
Shera, C. A., and Zweig, G. (**1993**). In *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 54–63.
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., and Tubis, A. (**1993**). In *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 25–32.
Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). "Modeling otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517–1543.
Talmadge, C. L., Tubis, A., Long, G. R., and Tong, C. (**2000**). "Modeling the combined effects of basilar membrane nonlinearity and roughness on stimulus frequency otoacoustic emission fine structure," J. Acoust. Soc. Am. **108**, 2911–2932.
van Dijk, P., and Manley, G. A. (**2009**). "The effect of ear canal pressure on spontaneous otoacoustic emissions: comparison between human and lizard ears," in *Concepts and Challenges in the Biophysics of Hearing*, edited by N. P. Cooper and D.T. Kemp (World Scientific, Singapore), pp. 196–201.
van Dijk, P., and Wit, H. P. (**1990**). "Synchronization of spontaneous otoacoustic emissions to a 2f1-f2 distortion product," J. Acoust. Soc. Am. **88**, 850–885.
van Dijk, P., and Wit, H. P. (**1998**). "Synchronization of cubic distortion spontaneous otoacoustic emissions," J. Acoust. Soc. Am. **104**, 591–594.
van Hengel, P. W. J., and Maat, A. (**1993**). In *Biophysics of Hair Cell Sensory Systems*, edited by H. Duifhuis, J. W. Horst, P. van Dijk, and S. M. van Netten (World Scientific, Singapore), pp. 47–53.
Verpy, E., Weil, D., Leibovici, M., Goodyear, R. J., Hamard, G., Houdon, C., Lefèvre, G. M., Hardelin, J. P., Richardson, G. P., Avan, P., and Petit, C. (**2008**). "Stereocilin-deficient mice reveal the origin of cochlear waveform distortions," Nature (London) **456**, 255–259.
Wable, J., and Collet, L. (**1994**). "Can synchronized otoacoustic emissions really be attributed to SOAEs?" Hear. Res. **80**, 141–145.
Wit, H. P. (**1986**). "Statistical properties of a strong spontaneous otoacoustic emission," in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. E. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, Berlin), pp. 221–228.
Zhang, T., and Penner, M. J. (**1998**). "A new method for the automated detection of spontaneous otoacoustic emissions embedded in noisy data," Hear. Res. **117**, 107–113.
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 evoked otoacoustic emissions," J. Acoust. Soc. Am. **98**, 2018–2047.
Zwicker, E., and Peisl, W. (**1990**). "Cochlear preprocessing in analog models, in digital models, and in human inner ear," Hear. Res. **44**, 209–216.
Zwicker, E., and Schloth, E. (**1984**). "Interrelation of different oto-acoustic emissions," J. Acoust. Soc. Am. **75**, 1148–1154.

Outer hair cell electromechanical properties in a nonlinear piezoelectric model

Yi-Wen Liu[a] and Stephen T. Neely
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 30 December 2008; revised 27 March 2009; accepted 27 May 2009)

A nonlinear piezoelectric circuit is proposed to model electromechanical properties of the outer hair cell (OHC) in mammalian cochleae. The circuit model predicts (a) that the nonlinear capacitance decreases as the stiffness of the load increases, and (b) that the axial compliance of the cell reaches a maximum at the same membrane potential for peak capacitance. The model was also designed to be integrated into macro-mechanical models to simulate cochlear wave propagation. Analytic expressions of the cochlear-partition shunt admittance and the wave propagation function are derived in terms of OHC electro-mechanical parameters. Small-signal analyses indicate that, to achieve cochlear amplification, (1) nonlinear capacitance must be sufficiently high and (2) the OHC receptor current must be sensitive to the velocity of the reticular lamina.
 [DOI: 10.1121/1.3158919]

PACS number(s): 43.64.Kc, 43.64.Ld [BLM] Pages: 751–761

I. INTRODUCTION

The outer hair cells (OHCs) in mammalian cochleae are thought to provide a feedback that results in cycle-by-cycle amplification of traveling waves (Patuzzi, 1996; Robles and Ruggero, 2001; Dallos, 2008). This amplification has been considered necessary to account for the fine tuning and high sensitivity of hearing (Shera, 2007). Two "active" mechanisms possibly contribute to this amplification: (1) a hair-bundle (HB) motility generated during mechano-electrical transduction (MET) (Hudspeth, 1997) and (2) a somatic motility generated by protein motors embedded in the lateral membrane (Brownell *et al.*, 1985; Ashmore, 1987). Both mechanisms are nonlinear: the force generated by the electromotile HB is a nonlinear function of HB deflection angle (Fettiplace *et al.*, 2006) and the contraction and the gating charge due to the electromotile membrane are both a nonlinear function of trans-membrane voltage (Santos-Sacchi, 1991).

Nonlinear micro-mechanical models have been constructed to describe HB motility (e.g., Tinevez *et al.*, 2006) and somatic motility (Iwasa and Adachi, 1997; Spector *et al.*, 1999), respectively. Recently, nonlinearity in HB motility has been incorporated in simulations of cochlear mechanics (Ramamoorthy *et al.*, 2007; Lu *et al.*, 2009). Nonlinearity in somatic motility, however, was ignored in these studies. This paper presents a step toward assessing the role of nonlinear OHC somatic motility in cochlear mechanics. A nonlinear piezoelectric membrane model is presented. The model captures various features in OHC somatic motility and can be integrated with cochlear models to simulate interaction between OHC and surrounding structures at auditory frequencies.

Mountain and Hubbard (1994) first proposed to model OHC somatic motility as a piezoelectric effect. Electromotility was represented by an ideal piezoelectric transformer. The transformer converts voltage to force and couples gating charge to axial contraction. The original model, being linear and one-dimensional, has been extended to describe the nonlinearity in electromotility (Iwasa and Adachi, 1997; Spector *et al.*, 1999) and the anisotropic stress-strain relations of the lateral wall (Tolomeo and Steele, 1995; Iwasa and Adachi, 1997). More recent theoretic extensions include a high-frequency resonance due to coupling between electrical and mechanical waves within the lateral wall (Weitzel *et al.*, 2003) and OHC volume non-conservation in slower processes (Allen and Fahey, 2006). Foundation of OHC piezoelectricity in statistical mechanics was also described (Iwasa, 2001). A comprehensive review of OHC membrane electromechanical models was given by Spector *et al.* (2006).

Piezoelectric membrane models have been incorporated into cochlea models (Lu *et al.*, 2006; Ramamoorthy *et al.*, 2007; Lu *et al.*, 2009) to simulate auditory tuning curves (Ruggero *et al.*, 1990; de Boer and Nuttall, 2000; Cooper, 1998; Ren and Nuttall, 2001). Using control theory, Lu *et al.* (2006) explained how an electrically low-pass OHC membrane would allow traveling waves to amplify at higher frequencies. For simplicity, all the electro-mechanical parameters were kept linear. Ramamoorthy *et al.* (2007) and Lu *et al.* (2009) introduced nonlinearity to the MET in their models. Both studies correctly simulated the broadening of cochlear tuning and the decrease in gain as intensity of the stimuli increases. However, the role of nonlinear somatic motility in cochlear tuning remains unexplored.

The two sources of motility have different positions in the OHC feedback loop. Relative to the OHC filter (Neely, 1985), HB motility must be placed before it, and somatic motility, after. Numerical simulation has shown that the spectral distribution of harmonic distortion is sensitive to the position of nonlinearity in the feedback loop (How *et al.*, 2009). Thus, it is conceivable that nonlinearities in HB motility and somatic motility may also have different contribution to other phenomena in cochlear signal processing.

[a]Author to whom correspondence should be addressed. Electronic mail: liuy@boystown.org

Therefore, it is desirable to model nonlinearities separately in the two sources of OHC motility so as to study macroscopic phenomena in cochlear mechanics.

This paper reports a step toward this goal. An OHC membrane model is proposed in Sec. II and an equivalent circuit diagram is shown. Nonlinear analyses of the circuit are performed, and results are compared to experimental data for nonlinear capacitance and voltage-dependent cell compliance. Section III demonstrates how the model can be integrated with a model of the cochlear partition, and tuning properties near the hearing threshold are derived via a small-signal analysis. In Sec. IV, OHC physiological conditions for traveling-wave amplification are further explored. The present model is compared to previous models, and discussions are given in Sec. V.

II. THE PROPOSED MEMBRANE MODEL AND ITS PREDICTIONS

In this section, an equivalent circuit model for the OHC membrane is presented. The circuit represents a relation between the OHC receptor current i_r and the membrane potential V. It predicts a stiffness-dependent nonlinear capacitance and a voltage-dependent compliance, and results are compared to experimental data.

A. Construction of a circuit diagram

First, let us assume that the OHC membrane has a dielectric capacitance C and a leakage conductance G. Thus, i_r can be written as the following:

$$i_r = GV + C\frac{dV}{dt} + i_d, \tag{1}$$

where i_d denotes a gating current due to charge displacement induced by conformational change in the membrane motors. Let $Q=\int i_d dt$ denote this charge and assume that Q is linearly coupled to OHC motility ξ_o:

$$\xi_o = TQ. \tag{2}$$

Here, a positive ξ_o represents a reduction in the length of OHC, and the coefficient T can be regarded as a piezoelectric transformer ratio (Mountain and Hubbard, 1994). Throughout this paper, T is defined as a constant independent of V. Further, assume that Q is a nonlinear function of V and an internal tensile force f_{OHC}:

$$Q = \frac{Q_{\max}}{1+\exp[-(V - Tf_{\mathrm{OHC}} - v_0)/v_1]}. \tag{3}$$

Here, $Q_{\max}$ is the maximum gating charge, and v_0 and v_1 determine the midpoint and the slope of $Q(V)$, respectively. Equation (3) is a consequence of a one-dimensional thermodynamic model and its derivation is given in the Appendix.

All the equations above can be summarized by the circuit diagram shown in Fig. 1, where $\dot{\xi}_o$ denotes the OHC contraction velocity, K denotes a static OHC axial stiffness, and Z represents a mechanical impedance if there is any external load.

This circuit is similar to that of Mountain and Hubbard (1994),[1] except that the gating current i_d charges a nonlinear capacitor in series with an ideal transformer. The circuit diagram is drawn this way to be consistent with the assumption that Q is a nonlinear function of $V-Tf_{\mathrm{OHC}}$ in Eq. (3).

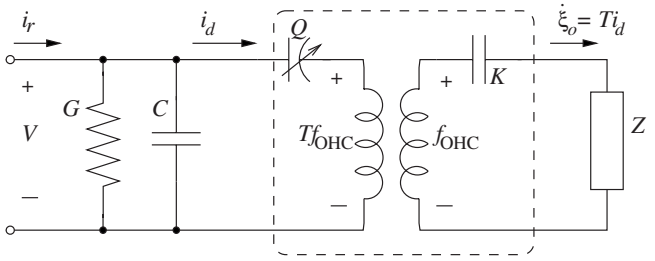


FIG. 1. The proposed circuit model of the OHC lateral membrane. The dashed box represents a nonlinear piezoelectric component. It is connected to the electrical domain on the left and the mechanical domain on the right (symbols are defined in the text).

It should be clarified that the piezoelectric component is always subject to the full membrane potential V. Nevertheless, it is useful to think of V as the sum of $V-Tf_{\mathrm{OHC}}$ and Tf_{OHC}. The advantage of such thinking will become more obvious when conducting small-signal analyses in Secs. III and IV.

At this point, an observation can be made: if the mechanical impedance of OHC is so small that $Tf_{\mathrm{OHC}} \ll V$, the membrane capacitance will be measured as the sum of a constant C and a nonlinear part $C_{\mathrm{NL}}=\partial Q/\partial V$. Since $Q(V)$ is a Boltzmann function in Eq. (3), its derivative $C_{\mathrm{NL}}(V)$ is bell-shaped. This agrees qualitatively with patch-clamp measurements in isolated OHCs (e.g., Santos-Sacchi, 1991). However, if the mechanical impedance Z is not negligible, the *nonlinear capacitance* C_{NL} can be computed as described next.

B. Nonlinear capacitance

First, let us assume that C_{NL} is measured by sweeping the voltage V slowly so that the mechanical impedance Z is stiffness-dominated. Denote the stiffness of Z as K_p. Thus, Hooke's law defines a relation between force f_{OHC} and displacement ξ_o as the following:

$$f_{\mathrm{OHC}} = (K + K_p)\xi_o. \tag{4}$$

Combining Eqs. (2)–(4), the following relation between Q and V is obtained:

$$Q = \frac{Q_{\max}}{1+\exp[-(V - T^2(K+K_p)Q - v_0)/v_1]}. \tag{5}$$

Note that the variable Q occurs on both sides of Eq. (5); the equation is transcendental and $Q(V)$ can only be solved numerically. Finally, C_{NL}, by definition, can be computed by taking the first derivative of Q with respect to V:

$$C_{\mathrm{NL}} \triangleq \frac{\partial Q}{\partial V}. \tag{6}$$

The function $Q(V)$ is plotted in Fig. 2(a) for three different values of stiffness: $K_p=0$ represents an OHC in isolation, $K_p=0.05$ N/m is typical of a fiber-glass probe contacting the OHC during a certain kind of measurement (e.g., Hallworth, 2007), and $K_p=0.15$ N/m is typical of OHC *in*

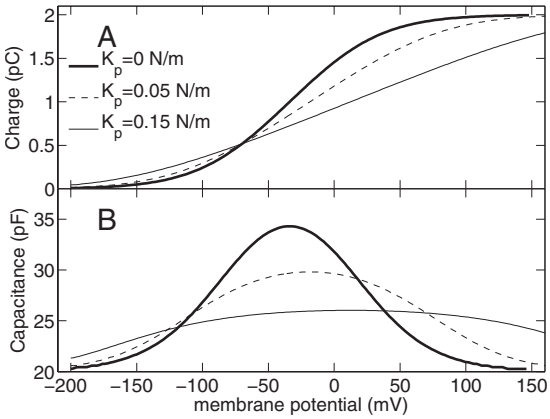


FIG. 2. (A) Gating charge and (B) the sum of C and C_{NL} as a function of membrane potential. Both functions are plotted using three different values of load stiffness K_p=0, 0.05, or 0.15 N/m. Other OHC parameters are K =0.02 N/m, $Q_{\max}$=2 pC, $T=8\times10^5$ m/C, v_0=−40 mV, and v_1=28.6 mV.

situ [see stiffness of reticular-lamina (RL), denoted as K_r in Table I]. Results show that, at a higher K_p, $Q(V)$ is more spread-out.

The sum of C and C_{NL} is shown in Fig. 2(b) as a function of V. Similar to $Q(V)$, it is more spread-out for a higher value of K_p. The peak capacitance decreases and the peak voltage shifts to the right as K_p increases.

In the extreme case when K_p approaches infinity,$C_{\mathrm{NL}}(V)$ should approach zero because the external load is too stiff for the OHC to make any contraction. This has been experimentally observed by Adachi and Iwasa (1999). However, the decrease in peak capacitance was not accompanied by a shift in peak voltage. This discrepancy can be resolved by considering mechanical orthotropy in a two-dimensional membrane model (Iwasa and Adachi, 1997). Here, the shift in peak voltage should be seen as an artifact due to one-dimensional approximation of membrane elasticity in the present model.

C. Voltage-dependence of cell compliance

There has been a dispute over the voltage-dependence of the axial stiffness of OHC. The axial stiffness decreased as a function of membrane potential in an experiment conducted by He and Dallos (1999). A thermodynamic model was constructed to explain this result (Deo and Grosh, 2004). However, no correlation was found between the cell compliance (inverse of stiffness) and membrane potential in a more recent experiment (Hallworth, 2007).

It would be an interesting exercise to see whether the present model predicts a voltage-dependence of the OHC axial compliance. First, by multiplying Eq. (3) by T, OHC contraction can be written as the following:

$$\xi_o = \frac{TQ_{\max}}{1+\exp[-(V-Tf_{\mathrm{OHC}}-v_0)/v_1]}. \tag{7}$$

Further, let us assume that a small external force f_{ext} is applied in order to measure the cell compliance. The external force reduces the internal tensile force f_{OHC} from its static value $f^{(0)}$; that is,

$$f_{\mathrm{OHC}} = f^{(0)} - f_{\mathrm{ext}}. \tag{8}$$

Consequently, the total contraction ξ_{tot} due to f_{ext} is given by the following:

TABLE I. List of parameters and their values used in simulation.

Symbol	Meaning (unit)	Value
	Organ of Corti mechanical parameters	
M_r	Mass of RL system (kg)	2.0×10^{-10}
K_r	Stiffness of RL system (N/m)	0.175
R_r	Damping of RL system (kg/s)	$(\sqrt{M_rK_r}/4.0)$[a]
M_b	Mass of BM system (kg)	2.0×10^{-10}
K_b	Stiffness of BM system (N/m)	1.2
R_b	Damping of BM system (kg/s)	$(\sqrt{M_bK_b}/6.0)$[a]
	Outer hair cell electro-mechanical properties	
T	Piezoelectric transformer ratio (m/C)	(8×10^5)[b]
G	Membrane conductance (nS)	50
C	Membrane capacitance (pF)	20
$\tilde{c}$	Small-signal equivalent capacitance (pF)	15
α_d	Receptor current's sensitivity to RL displacement (A/m)	(0.02)[c]
α_{v0}	Receptor current's sensitivity to RL velocity (C/m)	(2.5×10^{-6})[c]
K	Static axial stiffness (N/m)	(0.02)[d]
	Physical dimensions	
A	Cochlear cross-sectional area (cm^2)	0.01
W	BM width (cm)	0.1
D	Spacing of OHC in the wave-traveling direction (μm)	15

[a]Lu *et al.* (2006).
[b]Mountain and Hubbard (1994).
[c]See Sec. V C for comparison to an estimate from Fettiplace *et al.* (2006).
[d]Hallworth (2007).

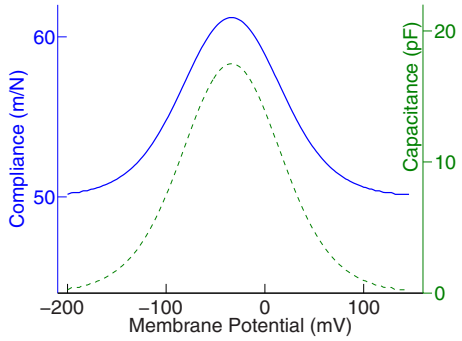


FIG. 3. (Color online) Cell compliance B (solid line) and small-signal equivalent capacitance $\tilde{c}$ (dashed line) as a function of membrane potential. Parameter values are $K=0.02$ N/m, $Q_{\max}=2$ pC, $v_0=-40$ mV, and $v_1=28.6$ mV.

$$\xi_{\text{tot}} = \frac{f_{\text{ext}}}{K} + \frac{TQ_{\max}}{1+\exp[-(V-Tf_{\text{OHC}}-v_0)/v_1]} = \frac{f_{\text{ext}}}{K} + \frac{TQ_{\max}}{1+\exp[-(V-Tf^{(0)}+Tf_{\text{ext}}-v_0)/v_1]},$$

and the cell compliance B can be calculated in the following way:

$$B \triangleq \frac{\partial \xi_{\text{tot}}}{\partial f_{\text{ext}}} = \frac{1}{K} + T\frac{\partial Q}{\partial \tilde{v}}\frac{\partial \tilde{v}}{\partial f_{\text{ext}}} = \frac{1}{K} + T^2\frac{\partial Q}{\partial \tilde{v}}.$$

In the preceding equation, $\tilde{v}$ denotes $V-Tf_{\text{OHC}}$ and can be regarded as an equivalent voltage across the nonlinear capacitor in Fig. 1. Therefore, $\partial Q/\partial \tilde{v}$ can be regarded as a small-signal equivalent capacitance of that capacitor $\tilde{c}$,

$$\tilde{c} \triangleq \frac{\partial Q}{\partial \tilde{v}} = \frac{Q_{\max}}{v_1}\cdot\frac{\theta}{(1+\theta)^2}, \tag{9}$$

where $\theta=\exp[-(V-Tf_{\text{OHC}}-v_0)/v_1]$. Consequently, B can be written as

$$B = \frac{1}{K} + T^2\tilde{c}. \tag{10}$$

It must be clarified that $\tilde{v}$ is a mathematical construct and it cannot be measured experimentally. The membrane motors are always subject to the full potential V and it is C_{NL} that is measured experimentally, not $\tilde{c}$.

Nevertheless, $\tilde{c}$ is a convenient term that will occur repeatedly in Secs. III and IV when analyzing OHC feedback and tuning. In Fig. 3, B and $\tilde{c}$ are shown as a function of V. According to Eq. (10), the cell compliance amount that can change with membrane potential is $T^2\tilde{c}(V)$. As shown in Fig. 3, the maximum compliance increment of 11 m/N corresponds to $\tilde{c}$ of 17 pF if $T=8\times10^5$ m/C.

This increment of 11 m/N in cell compliance corresponds to a fractional change of 0.22 for a hypothetical OHC with a static compliance K^{-1} of 50 m/N. The fractional compliance change reported by He and Dallos (1999) was about six to seven times greater than 0.22. Also, it increased monotonically as a function of membrane potential, whereas the present model predicts that the compliance reaches a maximum at approximately −40 mV. The present prediction is similar to a previous prediction by Iwasa [2001, Fig. 3(A)]. Implications of the discrepancy between the present model's prediction and He and Dallos's (1999) data are further discussed in Sec. V B.

Hallworth's (2007) data [2007, Fig. 3(c)], however, showed a smaller fractional compliance change in the range of −0.4 to +0.5. Most of the data points were scattered between ±0.25. Hallworth's (2007) fractional compliance change is comparable to what the present model predicts, but it did not have a single maximum as a function of voltage. Neither did it increase monotonically as a function of membrane potential. The fractional compliance change predicted by the present model, at most 0.22 if it exists, might be too small to have been observed by Hallworth (2007).

III. SMALL-SIGNAL ANALYSIS OF OHC FEEDBACK

Despite the partial success in explaining electromechanical properties experimentally, this section proceeds to integrate the present model with macro-mechanical cochlear models and calculate OHC feedback. Here and in Sec. IV, analyses will be conducted under the *small-signal* assumption; in other words, stimuli and responses are assumed to have small magnitudes so that the system is represented well by its linear approximation.

For the convenience of discussion, the OHC feedback is characterized by a transfer function $H_o(s)$ that relates OHC contraction to RL displacement,

$$H_o(s) = \xi_o(s)/\xi_r(s), \tag{11}$$

where $s=j\omega$ denotes the frequency variable in Laplace transform. $H_o(s)$ can be regarded as an open-loop displacement gain produced by the OHC. To calculate it, first, note that Eq. (2) relates ξ_o to i_d in the following way:

$$\xi_o = Ti_d/s, \tag{12}$$

where the factor $1/s$ represents integration with respect to time. Subsequently, i_d is related to i_r based on Kirchhoff's principle:

$$i_d = i_r\frac{(1/s\tilde{c}+Z_{\text{eq}})^{-1}}{(G+sC)+(1/s\tilde{c}+Z_{\text{eq}})^{-1}}. \tag{13}$$

In the preceding equation, $Z_{\text{eq}}=T^2(Z+s^{-1}K)$ denotes an equivalent electrical impedance due to mechanical load impedance Z and OHC static stiffness K. Finally, the receptor current i_r can be related to ξ_r by assuming that

$$i_r = (s\alpha_v+\alpha_d)\xi_r. \tag{14}$$

In the preceding equation, α_v and α_d represent a velocity-to-current and a displacement-to-current gain, respectively. Note that i_r responds to HB deflection in reality (e.g., Fettiplace *et al.*, 2006), but interaction between HB and the tectorial membrane (TM) is not modeled here in Eq. (14). Two different ways to interpret the equation will be given in Sec. V C.

Combining Eqs. (12)–(14), the transfer function $H_o(s)$ can be written as

$$H_o(s) = \frac{T(s\alpha_v+\alpha_d)/\tilde{c}}{s\tilde{c}+(G+sC)[1+T^2\tilde{c}(K+sZ)]}. \tag{15}$$

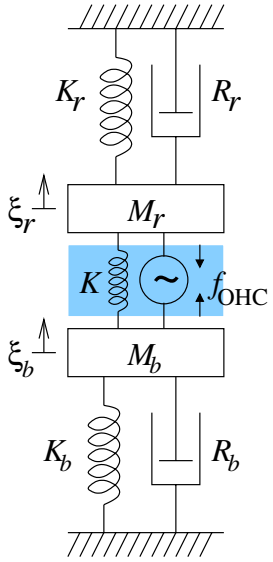


FIG. 4. (Color online) Organ of Corti micro-mechanical model (Lu *et al.*, 2006). The RL system is characterized by parameters $\{K_r, R_r, M_r\}$, the BM system is characterized by parameters $\{K_b, R_b, M_b\}$, and OHC lateral wall (shaded area) is characterized by a contraction force f_{OHC} and a static stiffness K.

The mechanical load impedance Z in Eq. (15) is not defined yet. As illustrated in Fig. 4, let us assume that the OHC contraction force f_{OHC} pulls two systems toward each other: a basilar-membrane (BM) system with an impedance of $Z_b \triangleq s^{-1}K_b + R_b + sM_b$ and a RL system with an impedance of $Z_r \triangleq s^{-1}K_r + R_r + sM_r$ (Lu *et al.*, 2006). Thus, Z can be written as

$$Z = \frac{Z_b Z_r}{Z_b + Z_r}. \tag{16}$$

Equation (16) can be substituted into Eq. (15) to calculate $H_o(s)$. Figures 5(a) and 5(b) show the magnitude and phase response of $H_o(s)$, respectively. The magnitude response of $H_o(j\omega)$ is plotted in logarithmic scale. Below 1 kHz, the magnitude response rolls off at 6 dB/octave due to the RC-filtering of the membrane represented by the term $(G+sC)$ in the denominator of Eq. (15). Between 1 and 4 kHz, the response is relatively flat because the 6 dB/octave roll-off is compensated by a 6 dB/octave gain due to the term $s\alpha_v$ in the numerator of Eq. (15). Between 4 and 16 kHz, the response has two resonance peaks due to the presence of two second-order systems: the RL and the BM. Above 16 kHz, the response rolls off at 12 dB/octave.

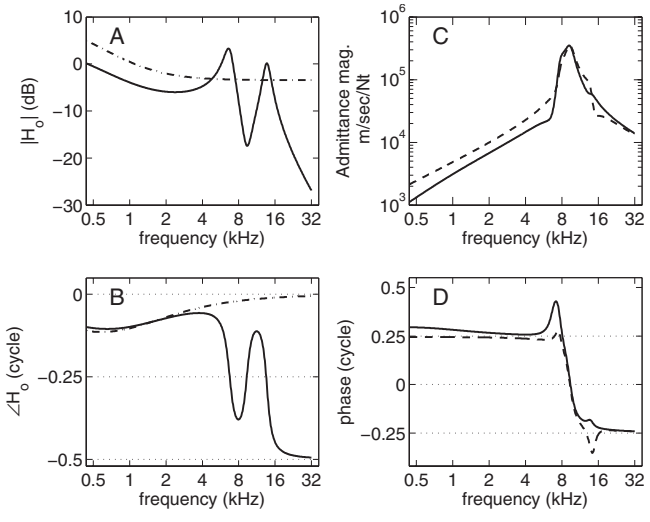


FIG. 5. OHC frequency responses *in situ*. [(A) and (B)] Magnitude and phase of $H_o(s)$. For comparison, dash-dotted curves show results for $Z=0$ (as in an isolated preparation). [(C) and (D)] Magnitude and phase responses, respectively, of velocities $s\xi_r$ (solid) and $s\xi_b$ (dashed) with respect to an externally applied force f_{cp}. Parameters used in this simulation are listed in Table I.

Although the term $s\alpha_v$ compensates for the 6 dB/octave loss of gain due to RC-filtering, there is no guarantee that an OHC gives amplification to cochlear traveling waves. Whether power is dissipated or amplified must be determined by the response of the cochlear partition to a force stimulus. Let us assume that the cochlear partition consisting of the BM system and the RL system is subject to an external force f_{cp}:

$$f_{cp} = Z_b \cdot s\xi_b + Z_r \cdot s\xi_r. \tag{17}$$

Though the preceding equation has two velocity variables $s\xi_b$ and $s\xi_r$, it has only one degree of freedom because the two variables are related in the following manner:

$$s\xi_b = s\xi_r + s\xi_o = [1 + H_o(s)] \cdot s\xi_r. \tag{18}$$

Combining Eqs. (17) and (18), the velocities can be written as a response to the force:

$$s\xi_r = \frac{1}{Z_b(1 + H_o) + Z_r} \cdot f_{cp}, \tag{19}$$

$$s\xi_b = \frac{1 + H_o}{Z_b(1 + H_o) + Z_r} \cdot f_{cp}. \tag{20}$$

The magnitude and phase response of these two velocities with respect to f_{cp} are shown in Figs. 5(c) and 5(d), respectively. The solid curves represent the RL response $s\xi_r/f_{cp}$, and the dashed curves represent the BM response $s\xi_b/f_{cp}$. These responses are referred to as "admittances" because their dimensionality is velocity divided by force. The magnitude responses increase then decrease as a function of frequency, indicating a compliance-dominance at low frequency and a mass-dominance at high frequency. Note that, near 6–8 kHz, the RL admittance phase is more than +0.25 cycle. This corresponds to negative damping, and Sec. IV examines it further.

IV. THE ROLE OF OHCs IN TRAVELING-WAVE AMPLIFICATION

The focus of this section is to calculate the damping coefficient for a one-dimensional, forward-traveling shear wave along the cochlear partition. The traveling of one-dimensional waves can be modeled as propagation in a transmission line, assuming that the canonical variables are pressure difference p across the cochlear partition and volume velocity U along the cochlea (Dallos, 1973, Chap. 4). Fluid

motion inside the organ of Corti, which can be handled by a three-chamber model (Lu *et al.*, 2006), is ignored under the transmission-line assumption.

A. Propagation function $k(s)$ in terms of OHC transfer function $H_o(s)$

At any single location along the cochlea, the damping coefficient $\gamma(s)$ is given as the real part of a propagation function $k(s)$ defined as the following;

$$k = \sqrt{z_{se} y_{sh}} \triangleq \gamma + j\kappa. \tag{21}$$

In the preceding equation, z_{se} and y_{sh} are an acoustic series impedance and an acoustic shunt admittance, respectively. The impedance z_{se} is defined as the pressure gradient $(-\partial_x p)$ divided by U; here, ∂_x denotes partial derivation along the direction of wave propagation. If viscosity of the cochlear fluid can be ignored, z_{se} is given by the following:

$$z_{sc} = s\frac{\rho}{A}, \tag{22}$$

where $\rho = 1.0$ g cm^{-3} is the density of cochlear fluid, and A is the cross-sectional area of the cochlea.

The shunt admittance y_{sh} is defined as the volume-velocity gradient $\partial_x U$ divided by $(-p)$. Let us assume that $\partial_x U$ equals the width W of cochlear partition times the sum of RL and BM velocities; that is,

$$\partial_x U(s) = W \cdot s(\xi_r + \xi_b). \tag{23}$$

For simplicity, also assume that p is a constant radially so that an effective force f_{cp} exerted on an OHC is given by the following;

$$f_{cp} = -p \cdot DW/3, \tag{24}$$

where D is the distance from one OHC to its nearest neighbor in the longitudinal direction, and the factor of 3 is the number of rows of OHCs radially. Combining the two preceding equations, y_{sh} can be written as the following:

$$y_{sh} \triangleq \frac{\partial_x U(s)}{-p(s)} = \frac{DW^2}{3} \cdot \frac{s(\xi_r + \xi_b)}{f_{cp}}. \tag{25}$$

Substituting Eqs. (19) and (20) into Eq. (25), the shunt admittance can be written in terms of the OHC transfer function:

$$y_{sh} = \frac{DW^2}{3} \cdot \frac{2 + H_o(s)}{Z_b(s)[1 + H_o(s)] + Z_r(s)}. \tag{26}$$

Finally, the propagation function $k(s)$ can be calculated by substituting Eqs. (22) and (26) into Eq. (21). The dimensionality of k is cm^{-1}. The real part γ is an attenuation factor, and the imaginary part κ is the wave number. If reverse-traveling waves can be ignored, Neely and Allen (2009) showed that the sign of γ determines whether the forward-traveling waves are attenuated $(\gamma > 0)$ or amplified $(\gamma < 0)$.

Because z_{se} is purely imaginary in Eq. (22), γ in Eq. (21) is negative only if the real part of y_{sh} is negative. Figure 6 shows the real and imaginary parts of y_{sh} and k given by Eqs. (26) and (21), respectively. Figures 6(a) and 6(c) suggest that significant negative damping occurs only when $\alpha_v > 0$. If

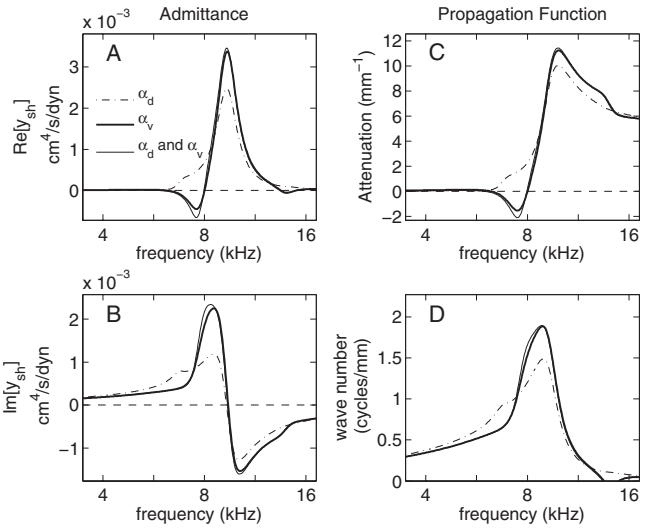


FIG. 6. Acoustic shunt admittance and propagation function as a function of frequency at a single location. [(A) and (B)] Real and imaginary parts of shunt admittance, respectively. [(C) and (D)] Real and imaginary parts of propagation function, respectively. Each panel shows results for three different motion-sensing conditions of the MET: $\alpha_v > 0$, $\alpha_d = 0$ (thick solid curves); $\alpha_v > 0$, $\alpha_d > 0$ (thin solid curves), and $\alpha_v = 0$, $\alpha_d > 0$ (dash-dotted curves).

$\alpha_v > 0$, a nonzero α_d provides a marginal improvement in the depth of negative damping. Also, negative damping occurs within half an octave below the characteristic frequency near 8 kHz, while the wave number κ reaches its maximum at a slightly higher frequency [Fig. 6(d)].

B. Variation in $y_{sh}(s)$ with respect to OHC parameters

The frequency response of shunt admittance $y_{sh}(s)$ is sensitive to OHC parameters. In this section, sensitivities to α_v and $\tilde{c}$ are studied because they represent HB motility and somatic motility, respectively. Figure 7 depicts variation in the real part of $y_{sh}(s)$ with respect to α_v and $\tilde{c}$. For $\tilde{c}$ of 2 pF, negative damping does not occur except for the two highest α_v values. More significant negative damping occurs for $\tilde{c}$ of 5 pF or higher and, as α_v increases, the depth of negative damping first increases (until $\alpha_v/\alpha_{v0} = 4$) and then decreases. For each $\tilde{c}$, the response shifts toward high frequency as α_v increases.

It is of practical interest to quantify the effect of negative damping by measuring its depth and width. The depth of negative damping can be defined as $(-\min \mathrm{Re}[y_{sh}])$, but the width of the negative damping region (NDR) needs to be defined more carefully so it conveys information about the effect size. Figure 8(d) illustrates a reasonable way of defining the width of NDR. The higher boundary of NDR is defined as the highest frequency at which $\mathrm{Re}[y_{sh}] \leq 0$. The lower boundary of NDR is defined via calculation of the slope of $\mathrm{Re}[y_{sh}]$ with respect to $\log f$: the tangential line at the point of maximum negative slope is illustrated by a thin line in Fig. 8(d), and its zero-crossing frequency defines the lower boundary of NDR.

In Figs. 8(a) and 8(b), the width and the depth of NDR are plotted as a function of $\tilde{c}$ for eight discrete values of α_v. For each α_v, negative damping only occurs if $\tilde{c}$ is higher than

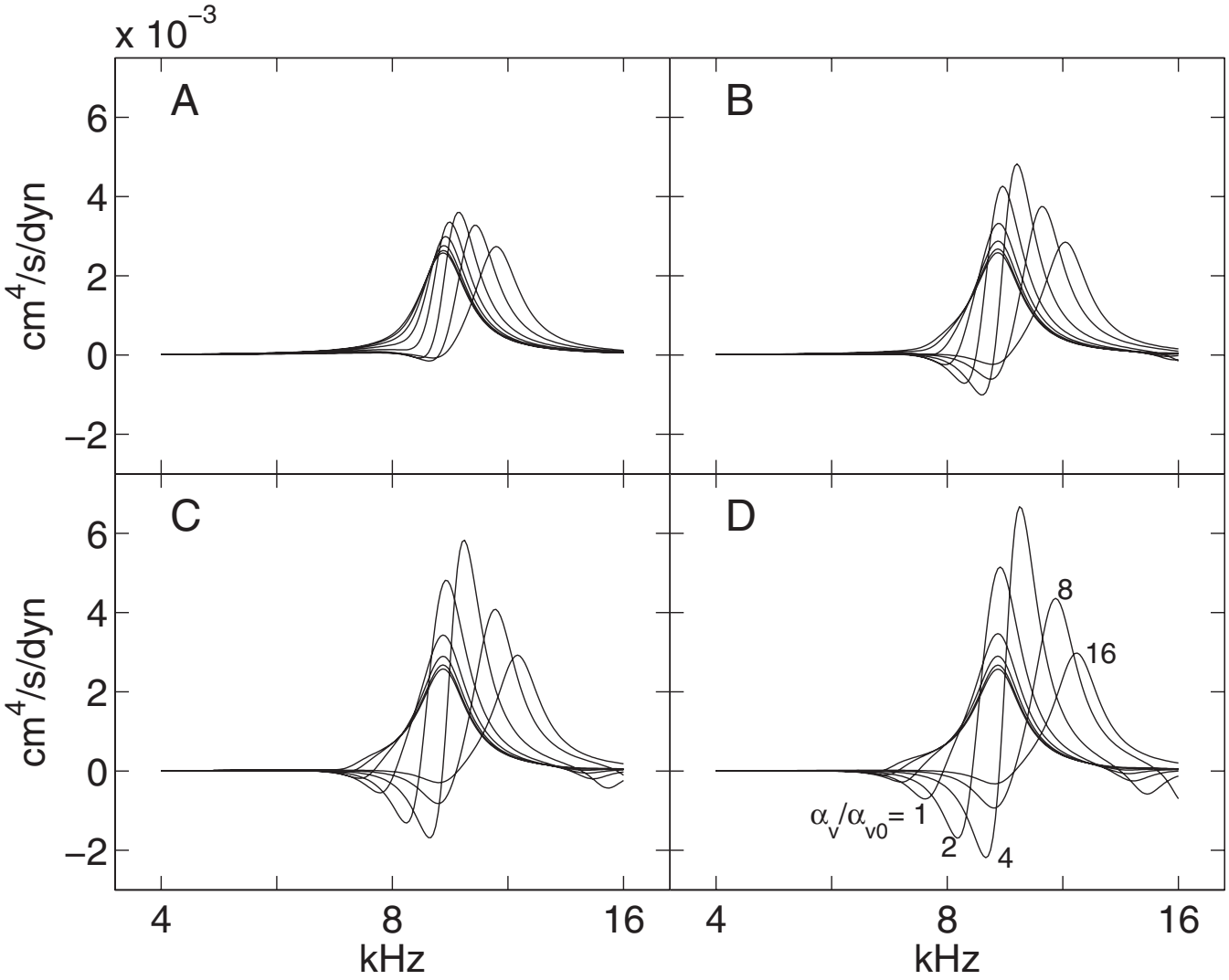


FIG. 7. The real part of shunt admittance y_{sh} as a function of frequency for different combinations of α_v and $\tilde{c}$. Panels (A), (B), (C), and (D) correspond to $\tilde{c}$=2, 5, 10, and 20 pF, respectively. In each panel, eight traces are plotted; from left to right, they correspond to α_v of 0.125, 0.25, 0.5, 1, 2, 4, 8, and 16 times the default $\alpha_{v0}=2.5\times10^{-6}$ C/m. Other parameters are listed in Table I.

a threshold. This threshold ranges from a few picofarads (for the highest α_v) to about 14 pF (for the lowest α_v). Figure 8(c) shows the product of width and depth of NDR. This product can be regarded as a figure of merit to evaluate how much negative damping an OHC provides to the traveling waves. As $\tilde{c}$ increases, the product reaches a plateau, and $\alpha_v=4\alpha_{v0}$ gives the highest plateau value among all choices of α_v.

C. Analytic approximation of shunt admittance $y_{sh}(s)$

Expressions for $y_{sh}(s)$ and $H_o(s)$ derived in Sec. IV A may be simplified to obtain analytic approximations of OHC tuning properties. To do so, note that in Fig. 7, NDR consistently occurs at higher than 6 kHz. At this frequency range, because $\omega \gg G/C$, conductance G can be neglected in Eq. (15). Also, α_d becomes negligible in Eq. (15) if α_v is sufficiently large so that

$$\alpha_v \gg \frac{\alpha_d}{\omega}. \quad \{\text{Assumption\#1}\}$$

Further, because $K_b \gg K_r$, if the frequency range of NDR is sufficiently lower than $\sqrt{K_b/M_b}$, it can be conveniently assumed that

$$|Z_b| \gg |Z_r|, \quad \{\text{Assumption\#2}\}$$

and thus $Z \approx Z_r$ in Eq. (16).

If Assumptions #1 and #2 are both valid, Eq. (15) has the following approximation:

$$H_o(s) \simeq \frac{\alpha_v/TC}{s^2M_r + sR_r + K + K_r + (T^2\tilde{c})^{-1} + (T^2C)^{-1}}. \quad (27)$$

By substituting Eq. (27) into Eq. (26), the following approximation of y_{sh} is obtained;

$$y_{sh}(s) \simeq \frac{DW^2}{3} \cdot \frac{2}{Z_b} \cdot \frac{s^2M_r + sR_r + K_{eq} + \alpha_v/2TC}{s^2M_r + sR_r + K_{eq} + \alpha_v/TC}, \quad (28)$$

where an equivalent stiffness K_{eq} is defined as

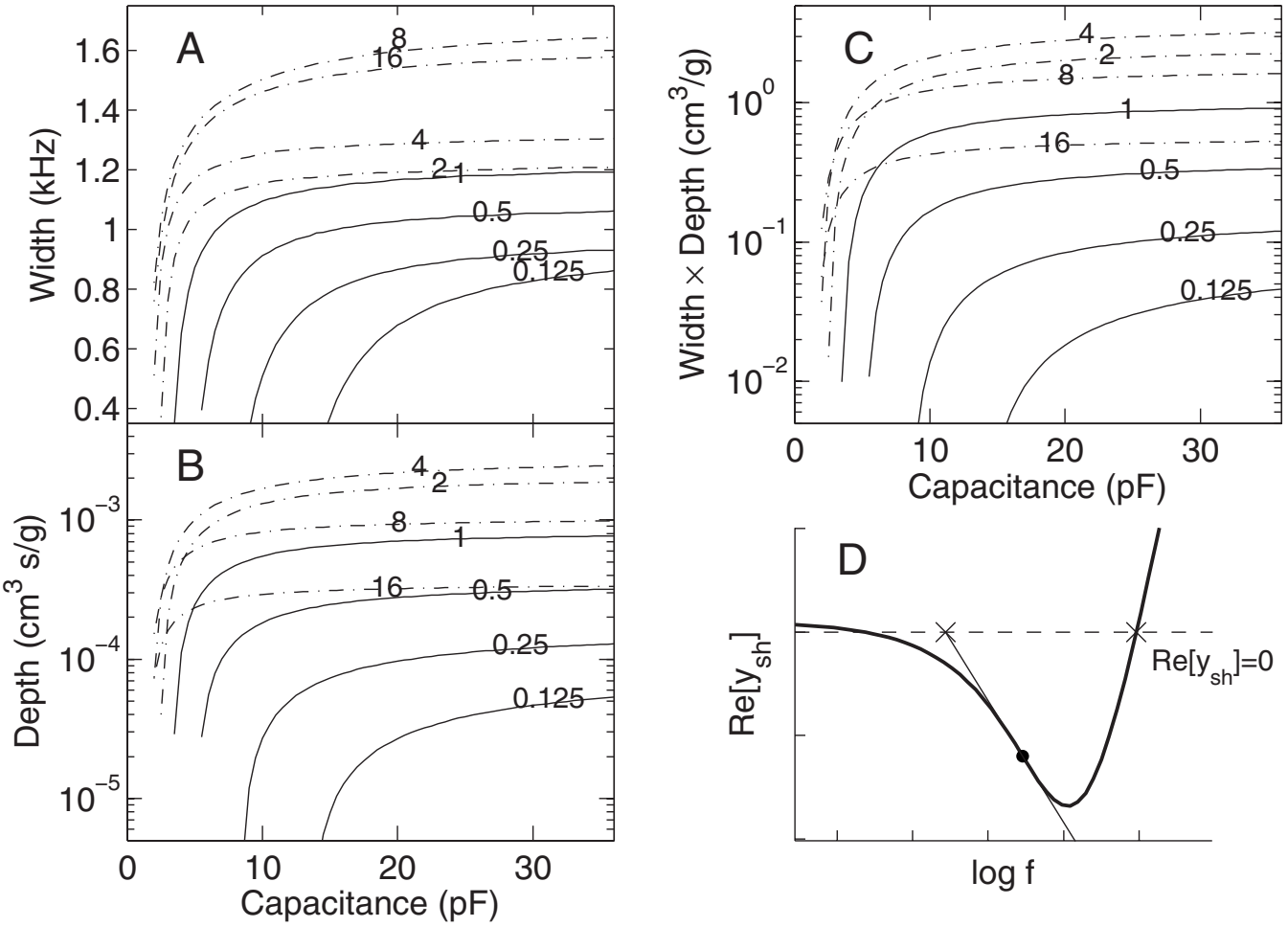


FIG. 8. Quantifying the effect of negative damping. [(A) and (B)] Width and depth, respectively, of NDR as a function of $\tilde{c}$. Each trace is marked by a number indicating the ratio α_v/α_{v0} (0.125, 0.25, etc.). For better viewing, results for $\alpha_v/\alpha_{v0}>1$ are plotted in dashed-dotted lines. Other parameters are listed in Table I. (C) Width times depth of NDR as a function of $\tilde{c}$. (D) Definition of NDR. The dashed line marks zero damping. The crosses mark the boundaries of NDR. The lower boundary is calculated by extending the tangential line from the point of maximum negative slope (marked by a dot).

$$K_{\text{eq}} = K + K_r + (T^2\tilde{c})^{-1} + (T^2 C)^{-1}.$$

In approximation (28), y_{sh} has a resonance frequency $\omega_b = \sqrt{K_b/M_b}$ due to Z_b, a pole frequency $\omega_p = \sqrt{(K_{\text{eq}} + \alpha_v/TC)/M_r}$, and a zero frequency $\omega_z = \sqrt{(K_{\text{eq}} + \alpha_v/2TC)/M_r}$. Note that ω_z is always lower than ω_p. If

$$\frac{K + K_r + 1/T^2\tilde{c} + (1 + T\alpha_v)/T^2 C}{M_r} < \frac{K_b}{M_b},$$

{Assumption#3}

in other words, if ω_p is lower than ω_b, then Z_b is stiffness-dominated at ω_p and y_{sh} has the following approximation:

$$y_{\text{sh}}(j\omega_p) \simeq \frac{DW^2/3}{K_b} \cdot \left(2j\omega_p - \frac{\alpha_v}{TR_r C}\right). \tag{29}$$

Although the approximation in Eq. (29) is rudimentary, it predicts that the real part of y_{sh} is negative at the pole frequency ω_p. This approximation can be refined if ω_p is still lower than but closer to ω_b:

$$y_{\text{sh}}(j\omega_p) \simeq \frac{DW^2/3}{K_b - \omega_p^2 M_b}\left(-\frac{\alpha_v}{TR_r C}(1-\chi) + 2j\omega_p(1+\chi)\right), \tag{30}$$

where

$$\chi = \frac{R_b}{K_b/\omega_p - \omega_p M_b} = \tan\left(\frac{\pi}{2} - \angle Z_b(j\omega_p)\right).$$

The real part of Eq. (30), being negative if $\chi<1$, gives an analytic approximation of the depth of NDR. Figure 9 compares this analytic prediction to numerical calculation of the depth of NDR described in Sec. IV B. The numerical results are obtained with the parameters listed in Table I, except that $\tilde{c}$ is set relatively high at 36 pF to ensure that NDR width reaches the plateau region in Fig. 8(b). Though it is derived under several assumptions, the analytic approximation produces less than 50% of error until α_v exceeds $4\alpha_{v0}$. For $\alpha_v = 8\alpha_{v0}$ and $16\alpha_{v0}$, the approximation fails because Assumptions #2 and #3 are not valid. Particularly, Assumption #3 requires that $\alpha_v < 5.5\alpha_{v0}$ under the present choice of parameter values. As α_v increases beyond, numerical calculation shows that NDR still exists but its depth decreases gradually as a function of α_v.

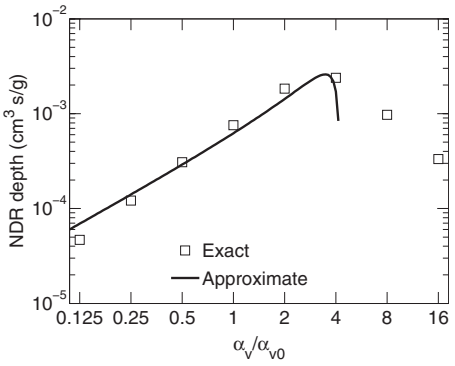


FIG. 9. Analytic approximation of negative damping depth as a function of α_v. The curve shows prediction of the depth given by Eq. (30) if $\tilde{c}$ approaches ∞. The squares show numerical calculation of the depth without simplifying the expression of $H_o(s)$ and $y_{sh}(s)$.

V. DISCUSSION

A. Comparison to other piezoelectric circuit models

Different circuit diagrams to represent OHC piezoelectricity have previously been drawn by Lu *et al.* (2006) and by Allen and Fahey (2006) independently. The model of Lu *et al.* (2006) was linear, and the displacement current i_d was proportional to the time-derivative of OHC external force. Transformers were not explicitly shown in the circuit diagram (Lu *et al.*, 2006, Fig. 18). In Allen and Fahey's (2006) model, it was suggested that nonlinearity be placed in the mechanical compliances so that nonlinear capacitance could be seen as a dual effect via piezoelectricity. In the present model, however, nonlinearity is placed in the capacitance, and nonlinear motility is seen as the dual effect.

The present circuit diagram may look at odds with Allen and Fahey's (2006) suggestion, but the symmetry of Gibbs energy with respect to voltage and force in Eq. (A3) implies that the circuit can be drawn either with a nonlinear compliance or a nonlinear capacitor. By substituting Q in Eq. (2) with Eq. (3), OHC motility ξ_o can be written as

$$\xi_o = \frac{L_{\max}}{1+\exp(f_{\mathrm{OHC}} - V/T - f_0)/f_1}, \tag{31}$$

where $L_{\max}=TQ_{\max}$, and f_0 and f_1 are defined in a similar manner as v_0 and v_1 in Eq. (3). Similar to $\tilde{v}$, a force component can be defined as $\tilde{f}=f_{\mathrm{OHC}}-V/T$. Thus, the circuit in Fig. 1 can be equivalently drawn as Fig. 10, where

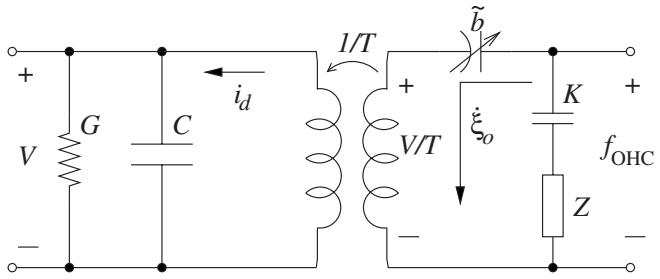


FIG. 10. An equivalent circuit to the proposed model in Fig. 1. The nonlinear component is moved from the electrical to the mechanical domain.

$$\tilde{b} \triangleq \frac{\partial \xi_o}{\partial \tilde{f}} \tag{32}$$

represents a nonlinear compliance. Note that $\tilde{b}$ is equal to $T^2\tilde{c}$ in Eq. (10). Also, when linearized, the circuit diagram in Fig. 10 becomes equivalent to what was presented by Lu *et al.* (2006, Fig. 18).

This circuit is similar to that of Allen and Fahey (2006, Fig. 2) in the sense that the force (or pressure in their case) is loaded with the series connection of a nonlinear compliance and a piezoelectric transformer. However, the model of Allen and Fahey (2006) was more complex because it allowed fluid to flow through the membrane. Such consideration may be important for modeling slower processes.

B. Voltage-dependent cell compliance

The present model cannot explain He and Dallos's (1999) finding of a monotonic decrease in stiffness as a function of membrane potential. Iwasa (2001) suggested that, to explain He and Dallos's (1999) results, the elastic moduli of the membrane motor must differ in its two conformational states. Without this refinement, Iwasa (2001) predicted a maximum axial-compliance increment of 20% as a function of membrane potential, which occurs at the same voltage for peak capacitance. Deo and Grosh (2004) carried out Iwasa's (2001) suggestion by assuming that the membrane motor is stiffer in the extended state than in the compact state and was able to reproduce the results of He and Dallos (1999) in a numerical simulation.

Compared to Iwasa's (2001) prediction, the present model produces similar results for the voltage-dependence of OHC axial-compliance. This is not unexpected because the present model can be derived from a one-dimensional approximation of Eq. (A2) proposed by Iwasa (1993).

Regarding Hallworth's (2007) findings, Dallos (2008) commented that only "healthier" cells demonstrate the voltage-dependent stiffness consistently. One possible explanation is that the membrane potential alters the self-association of motor molecules. A depolarizing (positive) potential works to "cluster" motor molecules and moves the equilibrium in favor of higher oligomeric forms rather than monomers (Rajagopalan *et al.*, 2007). This shift of equilibrium may also give rise to a stiffness decrease when the OHC is depolarized. A recent observation of decoupling between gating charge and motility in a simultaneous measurement (Wang, 2008) may provide more insight on this issue.

Whether or not OHC stiffness is voltage-dependent, cochlear tuning may not be sensitive to it. Note that the value of the axial stiffness K listed in Table I is typical of an OHC from the basal turn of the cochlea (Hallworth, 2007). This K is negligible in Eqs. (15) and (27) at the frequency range of interest (near NDR) because it is much smaller than K_r. Consequently, tuning should not be affected much by the voltage-dependence of OHC stiffness that was observed by He and Dallos (1999).

However, the scenario may be different elsewhere in the cochlea. Empirically, K is inversely proportional to the static length of OHC (Holley and Ashmore, 1988). Therefore, K

can be two to ten times smaller than the present value at more apical locations (He and Dallos, 1999) where OHCs are longer. In contrast, K_b and K_r decrease even more from base to apex by up to 100- to 600-fold near the helicotrema (Naidu and Mountain, 1998; Lu *et al.*, 2006). Because of a higher ratio K/K_r there, K is not negligible in Eq. (15). Consequently, the voltage-dependence of K, or the lack of it, could be an important factor in determining the tuning properties in the apical region of the cochlea.

C. Cochlear amplification: Is receptor current sensitive to RL velocity?

As the present model predicts, negative damping requires that the receptor current be sensitive to RL velocity. Similarly, Lu *et al.* (2009) suggested that a velocity-sensing HB accounts for more cochlear amplification than a displacement-sensing HB.

Experiments have shown that the receptor current has a fast adaptation to a step deflection in the HB (Fettiplace *et al.*, 2006). In other words, the current has a low-pass filtered response to a velocity impulse. By inspection, the velocity-sensing gain measured in the experiments was about 0.25 pA/μm s [Fettiplace *et al.*, 2006, Fig. 2(B)]. This estimate is conservative because the time-constant of adaptation was too short to be measured by the devices. Fettiplace *et al.* (2006) also suggested that the receptor current *in vivo* should be about four fold larger than measured in isolated OHC because of a higher endolymphatic potential and potassium as the major cation. This would partially eliminate the order-of-magnitude difference between their data and the default α_{v0} value listed in Table I. Nevertheless, one should remain cautious that the value of α_{v0} used in the present study may be unrealistically high.

In Eq. (14), a heuristic relation between the receptor current and the RL motion is described. A reasonable interpretation of the equation is that the TM is radially rigid, and HB deflection is directly proportional to RL displacement. Under this interpretation, Eq. (14) states that the receptor current is the sum of a response to HB deflection plus a fast adaptive component.

An alternative interpretation is that the receptor current responds to the angle of HB deflection only, but the current is effectively sensitive to RL velocity via the interaction between the HB and a non-rigid TM. Let ξ_t denote the displacement of TM and ζ denote the deflection angle of HB, which is proportional to $\xi_r - \xi_t$ (ξ_r being RL displacement). At high frequency when the mass of TM dominates, ξ_t is approximately zero and ζ is proportional to ξ_r. At a lower frequency, it is conceivable that the phase of ξ_t may lag behind that of ξ_r, giving rise to a phase lead of ζ relative to ξ_r. If this happens, there would be a component of HB deflection that is in-phase with RL velocity. Thus, cochlear amplification would not require fast adaptation of MET. However, details of TM mechanics are beyond the scope of the present study.

Evidence of cochlear traveling-wave amplification has been presented via solution of an inverse-scattering problem (Shera, 2007). The propagation function k was estimated from auditory-nerve recording data (e.g., Temchin *et al.*, 2005; van der Heijden and Joris, 2006), and results indicate that negative damping must have occurred at basal, middle, as well as apical locations in the cochlea (Shera, 2007, Fig. 8). The present prediction of k as a function of frequency [Figs. 6(c) and 6(d)] is similar to Shera's (2007) estimation inasmuch as, first, that negative damping occurs below the best frequency; second, the peak of $\kappa(\omega)$ occurs near the zero-crossing frequency of $\gamma(\omega)$; finally, the width of NDR is less than half of an octave [Fig. 8(b)]. These similarities suggest that the present OHC model may be useful for describing active wave propagation in the cochlea. Also, analyses described in Sec. IV predict the sensitivity of the width and depth of NDR to OHC parameters in specific ways. These predictions can thus be verified or refuted via experimental means.

ACKNOWLEDGMENTS

This study was supported by a grant from NIH-NIDCD (Contract No. R01-DC8318).

APPENDIX: FOUNDATION IN STATISTICAL MECHANICS

Piezoelectric models of the OHC (e.g., Mountain and Hubbard, 1994; Tolomeo and Steele, 1995) can be derived from thermodynamic principles if Gibbs free energy is known. The simplest form of Gibbs free energy was proposed by Iwasa (1993) under the following assumptions: (a) the membrane is embedded with motors that have one contracted state and one extended state that differ in area by $\Delta a < 0$ and (b) the change in area is coupled by a charge transfer $\Delta q > 0$. Thus, the probability P_s of any motor unit being in the contracted state is given by the Maxwell–Boltzmann distribution,

$$P_s = \frac{1}{1 + \exp\left[\dfrac{\Delta G}{k_B T_0}\right]}, \tag{A1}$$

where $k_B = 1.38 \times 10^{-23}$ J/deg is the Boltzmann constant, T_0 is the absolute temperature, and ΔG is the difference of Gibbs free energy between the two states. Iwasa (1993) showed that, if the surface tension on the membrane is isotropic, ΔG is given by the following equation:

$$\Delta G = \Delta G_0 - \Delta q \cdot V - \Delta a \cdot \tau, \tag{A2}$$

where ΔG_0 is a constant, V is the membrane potential, and τ denotes the surface tension on the membrane. Note that, in Eq. (A2), the term $(-\Delta a) \cdot \tau$ is the required mechanical energy for a motor to make a transition from the extended state to the contracted state.

In this study, the nonlinearity in the piezoelectric membrane is described by Eq. (3). The equation can be derived by considering a one-dimensional approximation of Eq. (A2). Imagine that all the motors are identical and aligned in the axial direction and can contract or stretch, as shown in Fig. 11. Assume that the contracted state and the extended state differ by $\Delta l < 0$ in length. Thus, the required mechanical energy to make a transition from the extended state to the contracted state can be written as $(-\Delta l) \cdot f_{\mathrm{OHC}}$, where f_{OHC}

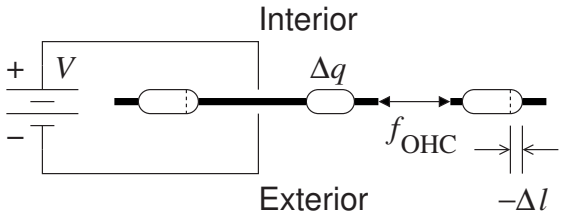


FIG. 11. Illustration of a one-dimensional model of piezoelectric motors embedded in the OHC lateral membrane. Of the three motors shown, two are in the extended state and the middle one is in the contracted state.

denotes the tensile force on the membrane. Hence, the difference in Gibbs energy between the two states is written as the following:

$$\Delta G = \Delta G_0 - \Delta q \cdot V - \Delta l \cdot f_{\mathrm{OHC}}. \tag{A3}$$

The total gating charge Q of an OHC that has N motor units on its lateral membrane is written as

$$Q = N\Delta q \cdot P_s. \tag{A4}$$

Combining Eqs. (A1), (A3), and (A4), Q can be written as a function of V and f_{OHC} in Eq. (3). The macroscopic parameters in Eq. (3) are related to the microscopic parameters in the following ways: $Q_{\max}=N\Delta q$, $T=-\Delta l/\Delta q$, $v_0=\Delta G_0/\Delta q$, and $v_1=k_{\mathrm{B}}T_0/\Delta q$.

[1]For simplicity, DC components such as the resting potential and the turgor pressure are not shown.

Adachi, M., and Iwasa, K. H. (**1999**). "Electrically driven motor in the outer hair cell: Effect of a mechanical constraint," Biophys. J. **96**, 7244–7249.

Allen, J. B., and Fahey, P. F. (**2006**). "Outer hair cell mechanics reformulated with acoustic variables," *Auditory Mechanisms: Processes and Models* (World Scientific, Singapore), pp. 194–201.

Ashmore, J. F. (**1987**). "A fast motile response in guinea-pig outer hair cells: The cellular basis of the cochlear amplifier," J. Physiol. (London) **388**, 323–347.

Brownell, W. E., Bader, C. R., Bertrand, D., and de Ribaupierre, Y. (**1985**). "Evoked mechanical responses of isolated cochlear hair cells," Science **227**, 194–196.

Cooper, N. P. (**1998**). "Harmonic distortion on the basilar membrane in the basal turn of the guinea-pig cochlea," J. Physiol. (London) **509**, 277–288.

Dallos, P. (**1973**). *The Auditory Periphery: Biophysics and Physiology* (Academic, New York).

Dallos, P. (**2008**). "Cochlear amplification, outer hair cells and prestin," Curr. Opin. Neurobiol. **18**, 370–376.

de Boer, E., and Nuttall, A. L. (**2000**). "The mechanical waveform of the basilar membrane. III. Intensity effects," J. Acoust. Soc. Am. **107**, 1497–1507.

Deo, N., and Grosh, K. (**2004**). "Two-state model for outer hair cell stiffness and motility," Biophys. J. **86**, 3519–3528.

Fettiplace, R., Crawford, A. C., and Kennedy, H. J. (**2006**). "Signal transformation by mechanotransducer channels of mammalian outer hair cells," *Auditory Mechanisms: Processes and Models* (World Scientific, Singapore), pp. 245–253.

Hallworth, R. (**2007**). "Absence of voltage-dependent compliance in high-frequency cochlear outer hair cells," J. Assoc. Res. Otolaryngol. **8**, 464–473.

He, D. Z. Z., and Dallos, P. (**1999**). "Somatic stiffness of cochlear outer hair cells is voltage-dependent," Proc. Natl. Acad. Sci. U.S.A. **96**, 8223–8228.

Holley, M. C., and Ashmore, J. F. (**1988**). "A cytoskeletal spring in cochlear outer hair cells," Nature (London) **335**, 635–637.

How, J., Elliott, S. J., and Lineton, B. (**2009**). "The influence on predicted harmonic generation of the position of the nonlinearity within micromechanical models," in *Concepts and Challenges in the Biophysics of Hearing*, edited by N. P. Cooper and D. T. Kemp (World Scientific, Singapore), pp. 350–351.

Hudspeth, A. J. (**1997**). "Mechanical amplification of stimuli by hair cells," Curr. Opin. Neurobiol. **7**, 480–486.

Iwasa, K. H. (**1993**). "Effect of stress on the membrane capacitance of the auditory outer hair cell," Biophys. J. **65**, 492–498.

Iwasa, K. H. (**2001**). "A two-state piezoelectric model for outer hair cell motility," Biophys. J. **81**, 2495–2506.

Iwasa, K. H., and Adachi, M. (**1997**). "Force generation in the outer hair cell of the cochlea," Biophys. J. **73**, 546–555.

Lu, S., Mountain, D., and Hubbard, A. (**2009**). "Is stereocilia velocity or displacement feedback used in the cochlear amplifier?," in *Concepts and Challenges in the Biophysics of Hearing*, edited by N. P. Cooper and D. T. Kemp (World Scientific, Singapore).

Lu, T. K., Zhak, S., Dallos, P., and Sarpeshkar, R. (**2006**). "Fast cochlear amplification with slow outer hair cells," Hear. Res. **214**, 45–67.

Mountain, D. C., and Hubbard, A. E. (**1994**). "A piezoelectric model of outer hair cell function," J. Acoust. Soc. Am. **95**, 350–354.

Naidu, R. C., and Mountain, D. C. (**1998**). "Measurements of the stiffness map challenge a basic tenet of cochlear theories," Hear. Res. **124**, 124–131.

Neely, S. T. (**1985**). "Mathematical modeling of cochlear mechanics," J. Acoust. Soc. Am. **78**, 345–352.

Neely, S. T., and Allen, J. B. (**2009**). "Retrograde waves in the cochlea," in *Concepts and Challenges in the Biophysics of Hearing*, edited by N. P. Cooper and D. T. Kemp (World Scientific, Singapore).

Patuzzi, R. (**1996**). "Cochlear micromechanics and macromechanics," in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 186–257.

Rajagopalan, L., Greeson, J. N., Xia, A., Liu, H., Sturm, A., Raphael, R. M., Davidson, A. L., Oghalai, J. S., Pereira, F. A., and Brownell, W. E. (**2007**). "Tuning of the outer hair cell motor by membrane cholesterol," J. Biol. Chem. **282**, 36659–36668.

Ramamoorthy, S., Deo, N. V., and Grosh, K. (**2007**). "A mechano-electro-acoustical model for the cochlea: Response to acoustic stimuli," J. Acoust. Soc. Am. **121**, 2758–2773.

Ren, T., and Nuttall, A. L. (**2001**). "Basilar membrane vibration in the basal turn of the sensitive gerbil cochlea," Hear. Res. **151**, 48–60.

Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.

Ruggero, M. A., Rich, N. C., Robles, L., and Shivapuja, B. G. (**1990**). "Middle-ear response in the chinchilla and its relationship to mechanics at the base of the cochlea," J. Acoust. Soc. Am. **87**, 1612–1629.

Santos-Sacchi, J. (**1991**). "Reversible inhibition of voltage-dependent outer hair cell motility and capacitance," J. Neurosci. **11**, 3096–3110.

Shera, C. A. (**2007**). "Laser amplification with a twist: Traveling-wave propagation and gain functions from throughout the cochlea," J. Acoust. Soc. Am. **122**, 2738–2758.

Spector, A. A., Brownell, W. E., and Popel, A. S. (**1999**). "Nonlinear active force generation by cochlear outer hair cell," J. Acoust. Soc. Am. **105**, 2414–2420.

Spector, A. A., Deo, N., Grosh, K., Ratnanather, J. T., and Raphael, R. M. (**2006**). "Electromechanical models of the outer hair cell composite membrane," J. Membr. Biol. **209**, 135–152.

Temchin, A. N., Recio-Spinoso, A., van Dijk, P., and Ruggero, M. A. (**2005**). "Wiener kernels of chinchilla auditory-nerve fibers: Verification using responses to tones, clicks, and noise and comparison with basilar-membrane vibrations," J. Neurophysiol. **93**, 3635–3648.

Tinevez, J.-Y., Jülicher, F., and Martin, P. (**2006**). "Unifying the various incarnations of active hair-bundle motility by the vertebrate hair cell," Biophys. J. **93**, 4053–4067.

Tolomeo, J. A., and Steele, C. R. (**1995**). "Orthotropic piezoelectric properties of the cochlear outer hair cell wall," J. Acoust. Soc. Am. **97**, 3006–3011.

van der Heijden, M., and Joris, P. X. (**2006**). "Panoramic measurements of the apex of the cochlea," J. Neurosci. **26**, 11462–11473.

Wang, X. (**2008**). "Outer hair cell electromotility and cochlear amplifier," Ph.D. thesis, Creighton University, Omaha, NE.

Weitzel, E. K., Tasker, R., and Brownell, W. E. (**2003**). "Outer hair cell piezoelectricity: Frequency response enhancement and resonance behavior," J. Acoust. Soc. Am. **114**, 1462–1466.

Voice gender perception by cochlear implantees[a)]

Damir Kovačić[b)]
Cognitive Neuroscience Sector, Scuola Internazionale Superiore di Studi Avanzati (SISSA), Via Lionello Stock 2/2, 34135 Trieste, Italy and SUVAG Polyclinic, Kneza LJ. Posavskog 10, 10000 Zagreb, Croatia

Evan Balaban
Behavioral Neurosciences Program, Stewart Biological Sciences Building, McGill University, 1205 Avenue Dr. Penfield, Montreal, Quebec H3A 1B1, Canada and Laboratorio de Imagen Médica, Hospital General Universitario Gregorio Marañón, Calle Dr. Esquerdo 46, 28007 Madrid, Spain

(Received 5 June 2008; revised 4 March 2009; accepted 26 May 2009)

Gender identification of human voices was studied in a juvenile population of cochlear implant (CI) users exposed to naturalistic speech stimuli from 20 male and 20 female speakers using two different voice gender perception tasks. Stimulus output patterns were recorded from each individual CI for each stimulus, and features related to voice fundamental frequency and spectral envelope were extracted from these electrical output signals to evaluate the relationship between implant output and behavioral performance. In spite of the fact that temporal and place cues of similar quality were produced by all CI devices, only about half of the subjects were able to label male and female voices correctly. Participants showed evidence of using available temporal cues, but showed no evidence of using place cues. The implants produced a consistent and novel cue to voice gender that participants did not appear to utilize. A subgroup of participants could discriminate male and female voices when two contrasting voices were presented in succession, but were unable to identify gender when voices were singly presented. It is suggested that the nature of long-term auditory categorical memories needs to be studied in more detail in these individuals.
 [DOI: 10.1121/1.3158855]

PACS number(s): 43.64.Me, 43.71.Bp, 43.66.Ts [RYL] Pages: 762–775

I. INTRODUCTION

Cochlear implant (CI) users exhibit high variability in their ability to perceive speech (Svirsky *et al.*, 2000; Blamey *et al.*, 2001). An important goal of CI research is to elucidate potential factors underlying this variation. Anecdotal evidence and previous studies (Cleary and Pisoni, 2002; Fu *et al.*, 2004; Spahr and Dorman, 2004; Cleary *et al.*, 2005; Fu *et al.*, 2005) suggest that even seemingly simple tasks such as identifying speakers pose challenges for CI users.

The identification of voice gender must be ultimately based on the quality of the spectral and temporal cues that speech items provide, even though individual variation in each individual's memory, experience, and motivation are also necessarily involved. Gender-related features of vocal tract anatomy [via body size effects on vocal tract length (VTL)], and laryngeal fold size [influencing differences in the temporal rate of vibration, which are reflected in differences in fundamental frequency (F0)] provide normal-hearing (NH) listeners with cues for voice gender identification. An almost perfect recognition of voice gender is achieved in NH individuals when both VTL and F0 cues are used (Bachorowski and Owren, 1999; Owren *et al.*, 2007; Smith *et al.*, 2007). In addition, automated recognition systems using VTL and F0 show robust performance in the face of within- and between-speaker acoustic variations (Childers and Wu, 1991; Wu and Childers, 1991).

CIs appear to elicit temporal pitch percepts only below ~300 Hz (Carlyon and Deeks, 2002; Zeng, 2002) and have inherently low spectral resolution (Cohen *et al.*, 1996; McKay, 2005). CI users may rely on particular isolated spectral or temporal cues, or some combination in gender identification tasks (Fu *et al.*, 2004; Laneau and Wouters, 2004b; Fu *et al.*, 2005; Chang and Fu, 2006). Temporal cues are delivered in the form of envelope modulations to one or more stimulation electrodes, while spectral cues are provided via the spatial pattern of electrode array stimulation (McKay *et al.*, 1994, 1995; Laneau and Wouters, 2004a, 2004b). Such cues should yield high levels of gender identification when voices are far apart from each other in F0 and VTL, even though CI users are poor at distinguishing speakers with similar vocal characteristics (Cleary and Pisoni, 2002; Fu *et al.*, 2005). However, the extent to which the implants of different users provide information of similar quality about natural speech signals may also play a role in individual variation in CI listener performance.

The present study examines CI users' perception of voice gender by assessing the degree to which individual CIs reliably transmit potentially available voice gender cues to their wearers, and the dependence of behavioral performance on the quality of this information.

[a)]Portions of these data were presented at the Conference for Implantable Auditory Prosthesis 2007, Lake Tahoe, CA.

[b)]Author to whom correspondence should be addressed. Present address: Laboratory of Auditory Neurophysiology, K.U. Leuven, Campus Gasthuisberg O&N 2, Herestraat 49 bus 1021, B-3000 Leuven, Belgium. Electronic mail: Damir.Kovacic@med.kuleuven.be

A. Availability of temporal cues to voice gender

Previous studies examining signals with amplitude modulated envelopes, both in hearing subjects (Burns and Viemeister, 1976, 1981) and CI users (McKay *et al.*, 1995; McKay and McDermott, 2000), show that the fundamental frequency of the input signal can be effectively encoded by temporal modulations of the signal envelope. Within the signal processor of each implant, input signals are processed into frequency sub-bands by means of a set of bandpass filters. Each sub-band is then assigned to the most relevant electrode in such a way that tonotopic organization is roughly preserved. Sub-bands with high-frequency content are delivered to the most basal electrodes, and sub-bands with low-frequency information are delivered to electrodes closer to the cochlear apex. In order to minimize perception of the carrier frequency and to correctly represent envelope modulations related to F0, the carrier frequency is usually at least four times higher than the typical F0 value (McDermott and McKay, 1994).

To find channels that are likely to represent F0 in their envelopes, the modulation spectrum (MS) of the stimulus sounds was examined after being processed by the same filter-bank settings used in each subject's device. By comparing the acoustic MS (calculated from bandpass-filtered acoustic signals) with MSs obtained from electrodograms captured from a subject's device, the electrode channel(s) that are theoretically able to deliver F0-related temporal cues relevant for gender identification in each subject could be identified. These channels are said to have F0 modulation availability, and were used to ascertain the relationship between perceptual performance and the quality of available F0 information for the different stimuli.

B. Availability of place cues to voice gender

Previous research with both CI devices and CI simulations delivered to NH subjects suggests that place cues for voice gender identification are available to CI users (Fu *et al.*, 2004, 2005; Gonzalez and Oliver, 2005). Laneau and Wouters (2004a) also demonstrated possible utilization of place cues for pitch discrimination in multi-electrode stimulation settings, with the average just-noticeable differences for place pitch ranging from 0.25 to 0.46 mm.

The present study assessed the availability of place cues for voice gender based on the relationship between stimulated electrode location and the vocal pitch or spectral envelope distribution of the speakers. Each stimulus produced a specific spatial output profile of electrode stimulation, quantified here as the cumulative sum of pulses in each electrode during the 2 s of stimulation provided by each stimulus. A median central location for this electrode stimulation pattern, equivalent to the "center-of-mass" of the area below the curve in the stimulation output profile, was calculated. The correlation between median electrode values and either the acoustic F0 of speech items or the distribution of their spectral envelope energy was then examined.[1]

C. Overall design of the experiments

Quantitative measures of temporal and place information that the implants provide about vocal stimuli were related to variation in behavioral responses to these same stimuli using two different testing procedures. The first was a fixed identification procedure with a one-interval, two-alternative forced choice (2AFC) task in which subjects indicated whether the speech sample was spoken by a male or female speaker. The second was an adaptive two-interval 2AFC discrimination procedure in which the adaptive parameter was the difference in the fundamental frequencies (F0) of male and female speech items (subjects responded by indicating which speech item was uttered by a female speaker). These two procedures were employed in order to assess the potential role of long-term auditory representations in CI users' performance. Allowing direct, short-term comparisons of two speech samples permitted additional evaluation of the auditory abilities of subjects who were not able to distinguish voice gender in the fixed procedure. The adaptive procedure also directly assessed how voice gender identification abilities relate to the size of F0 differences. The two procedures employed different speech items spoken by the same set of 20 male and 20 female speakers. This number of speakers was used to ensure that general processing strategies for the differences between male and female voices were being studied, rather than specific memorization strategies that could be used to discriminate among a small number of voices.

II. METHODS

A. Subjects

The study was approved by the ethical committees of the School of Medicine, University of Zagreb, Polyclinic SUVAG, and the Croatian Medical Chamber. Forty-one CI subjects with devices manufactured by Cochlear Corporation (20 males and 21 females; age range: 5.3–18.8 years, mean age=12.3 years) were recruited into the study using a database maintained by the Polyclinic SUVAG (a Croatian national institute for the rehabilitation of listening and speech). Subject details are given in Table I. One purpose of this research was to establish the relationship between the signal information being delivered to each individual CI user and their perceptual behavior. Since the apparatus for capturing stimulus output patterns was only provided by one CI manufacturer (Cochlear Ltd.), and was not compatible with devices of other manufacturers, it was necessary to limit the subject pool to individuals using Cochlear Ltd. devices (Esprit 3G and Sprint).

Psychological assessments of the participants were consulted to disqualify any subjects with reduced cognitive abilities that might influence their gender identification performance (Waltzman *et al.*, 2000; Holt and Kirk, 2005; Stacey *et al.*, 2006). Nonverbal psychological assessment included at least two or more of the following tests: Raven progressive matrices, Goodenough IQ, nonverbal WISC IQ, Brunet–Lezine, and Leiter International Performance Scale. These tests were performed by professional psychologists at least once during the subject's association with the Polyclinic

TABLE I. Characteristics of the CI participant population. Asterisks (*) denote unsuccessful capture of CI electrical signals.

Subject	Age at testing (years; months)	Sex	Ear	Processor type	No. of electrodes	Stimulation rate per channel (Hz)	Sound coding strategy
CI01	5; 4	F	R	Esprit 3G	20	1200	ACE
CI02	10; 6	M	R	Sprint	22	1200	ACE
CI03	12; 1	M	L	Esprit 3G	20	900	ACE
CI04	15; 8	F	R	Esprit 3G	20	900	ACE
CI05	18; 9	M	L	Esprit 3G	20	900	ACE
CI06	11; 4	M	R	Sprint	21	1200	ACE
CI07	12; 2	M	R	Esprit 3G	20	900	ACE
CI08	6; 9	M	L	Esprit 3G	20	1200	ACE
CI09	9; 5	M	R	Esprit 3G	20	1200	ACE
CI10	11; 0	F	L	Esprit 3G	20	900	ACE
CI11	11; 1	M	R	Esprit 3G	20	900	ACE
CI12	14; 0	M	R	Esprit 3G	20	900	ACE
CI13*	14; 7	F	R	Esprit 3G	20	250	SPEAK
CI14	11; 3	M	L	Esprit 3G	20	900	ACE
CI15	17; 6	F	R	Esprit 3G	20	900	ACE
CI16	7; 9	M	R	Esprit 3G	20	1200	ACE
CI17	14; 5	F	R	Esprit 3G	20	1200	ACE
CI18	13; 1	F	R	Esprit 3G	20	900	ACE
CI19*	14; 4	F	R	Esprit 3G	18	500	ACE
CI20*	9; 7	F	R	Esprit 3G	20	900	ACE
CI21	11; 0	M	R	Sprint	19	1200	ACE
CI22	18; 5	F	L	Esprit 3G	20	900	ACE
CI23	9; 2	F	R	Esprit 3G	20	1200	ACE
CI24	14; 9	F	R	Esprit 3G	20	900	ACE
CI25	8; 8	F	R	Esprit 3G	19	1200	ACE
CI26	17; 3	M	R	Esprit 3G	20	900	ACE
CI27*	12; 9	F	R	Esprit 3G	16	900	ACE
CI28*	8; 5	F	L	Esprit 3G	19	1200	ACE
CI29	11; 4	M	R	Esprit 3G	20	1200	ACE
CI30	8; 9	M	R	Sprint	22	1200	ACE
CI31	15; 7	M	L	Esprit 3G	20	1200	ACE
CI32	13; 11	F	R	Esprit 3G	20	900	ACE
CI33	14; 0	F	R	Esprit 3G	20	900	ACE
CI34	12; 11	F	R	Esprit 3G	20	900	ACE
CI35*	14; 7	F	R	Esprit 3G	19	720	ACE
CI36	10; 5	F	R	Esprit 3G	20	1200	ACE
CI37	15; 10	M	R	Esprit 3G	20	900	ACE
CI38	10; 11	M	R	Esprit 3G	19	1200	ACE
CI39	9; 1	M	R	Sprint	22	1200	ACE
CI40	14; 4	M	L	Esprit 3G	20	900	ACE
CI41	10; 0	F	L	Sprint	22	900	ACE
Mean	12.3						
SD	3.2						

SUVAG as a part of their hearing rehabilitation care. None of the subjects who participated in the present study had significant nonverbal psychological disorders. Spoken language proficiency was compromised in some cases because of previous auditory deprivation (Svirsky *et al.*, 2000, 2004). However, no minimal requirements in language proficiency were used as part of the selection criteria for the present study. Both the simplicity of the gender identification task (not requiring complex linguistic skills), and the fact that every participant in the study was able to provide their own explanation of what they were supposed to do, suggest that variations in spoken language proficiency did not create problems for subjects understanding the nature of the experimental tasks.

Control participants were 15 hearing children (8 males and 7 females, age range 6.7–10.6 years, mean age 9.3 years) recruited from one primary school in Zagreb. Neither the participants and their parents, nor their teachers reported hearing problems. In each case, parents or caregivers signed a consent form before their children participated in this study. The sex composition did not differ between the control (C) and experimental (E) group ($\chi^2=0.09$, $p>0.99$), but

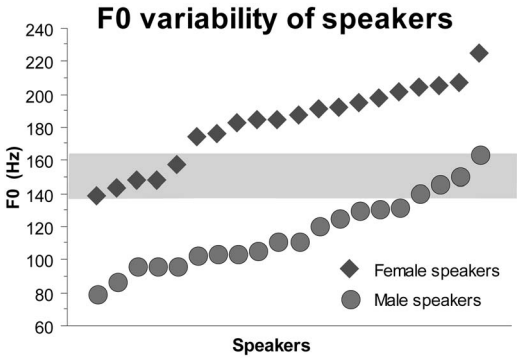


FIG. 1. Fundamental frequency (F0) values for the 20 male and 20 female speakers used in this study. The mean female F0 = 183.3 ± 5.4 Hz, and the mean male F0 = 117.9 ± 4.8 Hz. The shaded horizontal band denotes the region of overlapping fundamental frequencies (between 137.9 and 163.3 Hz). With overlapping speakers excluded, the mean female F0 = 195.1 ± 3.1 Hz and the mean male F0 = 110.1 ± 3.3 Hz.

the age composition differed (Mann–Whitney U-test, $\text{mean}_C = 9.3$, [SD = 1.3] years, $\text{mean}_E = 12.3$, [SD = 3.2] years, $n_C = 15$, $n_E = 41$, $p < 0.001$). The primary goal of the control group was to check whether the stimuli were of abnormal difficulty; due to the reduced hearing experience of CI subjects and their delays in language acquisition (Svirsky *et al.*, 2000, 2004; Nicholas and Geers, 2007), it is appropriate to compare older CI subjects with younger hearing controls.

B. Stimuli

Speech samples in the form of short news-like stories from 20 different male and 20 different female professional radio announcers were obtained from the national broadcasting radio company *Hrvatski radio* in digitized format, sampled at 44100 Hz using a 16-bit coding scheme. The samples were cut into 40 speech items of 2-s length with the following two requirements: (i) the onset was always aligned with the word onset and (ii) the offsets were never within phoneme boundaries. Since it was not possible to maintain the exact length of 2000 ms for all speech items in this way, the PSOLA lengthening algorithm (Moulines and Laroche, 1995) was used to equalize the length of all utterances. This manipulation was done using PRAAT software (Boersma and Weenink, 2006) with scaling factors between 0.84 and 1.25. At these scaling values, the PSOLA lengthening algorithm preserves pitch contours. Fundamental frequencies of the speakers were calculated using the autocorrelation method (Boersma, 1993) implemented in PRAAT software. The average F0 values of the speech items are shown in Fig. 1. Natural F0 variability occurs in this stimulus set both within the utterances of each speaker (pitch contours not shown) and between speakers. The male and female stimulus populations differed in their average fundamental frequencies [average female F0 = 183.3 ± 5.4 Hz and average male F0 = 117.9 ± 4.8 Hz, about 56% of one octave apart ($Z = 5.13$, $n = 20$, $p < 0.0001$, Mann–Whitney U-test)]. Natural variability in this sample yielded an overlapping area of F0 values between five female and four male voices, as indicated in Fig. 1 by the horizontal band. This overlapping range was between 137.9 (minimal female F0) and 163.3 Hz (maximal male F0). With overlapping speakers removed, the mean female F0 was 195.1 ± 3.1 Hz and the mean male F0 was 110.1 ± 3.3 Hz.

C. Voice gender identification

Prior to the commencement of the experiment, each subject was given instructions and performed practice trials until the experimenter was assured that the subject understood the task. The practice and experimental trials used PRESENTATION software (Version 10.1, Neurobehavioral Systems, Inc., Albany, CA). In each trial, one 2-s-long speech item was chosen randomly, and delivered to either headphones (hearing control subjects) or to the CI device (CI subjects) via direct line input. The subject was requested to respond by clicking on one of two response buttons representing a male and a female. The response buttons were associated with a computer display showing sketches of typical male and female faces. Feedback (a smiling face for correct responses and the symbol "X" for incorrect responses) was given in order to maintain the subject's interest in the experiment. The items were not replayed, even for incorrect responses. The first and second halves of the trials had response button positions exchanged (in the first half, the female was associated with the left button). Five sets of randomly chosen speech items were created and their presentation was counterbalanced between subjects. Practice consisted of six trials with two different female and male speech items from the same speaker database that were not used in the experimental trials. Response buttons (indicating male or female voice) were interchanged after half of the practice trials. CI subjects typically needed just one practice run before understanding the task and starting the experimental phase. To make sure that the participant understood the task, they were instructed to give their own explanation of what they had to do to the experimenter. Since the aim of this experiment was to assess each participant's ability to identify the gender of a voice, their performance did not have to reach a pre-defined criterion in order to progress to the experimental phase. Both the simplicity of the task, and the visual reinforcement with sketched pictures of male and female faces during response periods allowed all subjects (including two 4-year-old hearing subjects in a pilot study) to perform the task reliably throughout the practice trials and the experiment.

For the hearing control subjects, Sennheiser HD 580 headphones were used, with monaural presentation of stimuli to the right ear at 65 dB sound pressure level (SPL) (*A*-level) as measured by a RadioShack sound level meter (model 33-2055). Except for this difference in the stimulus delivery apparatus, the items and procedures were the same in both hearing control and CI subjects. After completion of the fixed identification procedure, the subjects had a short break (10 min) and then resumed their participation by undergoing a second procedure, described below. When the second adaptive procedure finished, the electrode stimulation patterns generated by all of the presented speech items were captured using processor control interface (PCI) and interface card (IF5) hardware (manufactured by Cochlear Ltd., Sydney, Australia and provided by Cochlear AG, Basel, Switzerland)

and the RFSTAT software system [manufactured and provided by the Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation (CRC HEAR), Melbourne, Australia]. CI signal capture was performed by recording two long audio sequences played to the subject's CI device with the transmitting coil attached to the PCI+IF5 system (instead of the subject's implanted receiver coil), keeping the same device settings used during the experiment. Each audio sequence was 84.25 s long and contained all speech items presented in the two procedures. Speech items were sequentially placed one after the other in the audio sequence and were separated by 100 ms of silence. In order to synchronize the capture signal onset with that of the audio signal, a 50-ms white noise burst was placed at the beginning of each long sequence. The analysis of CI captured signals was performed using custom software written in MATLAB (The Mathworks, Natick, MA). While the signals were being captured, subjects were led to an adjoining room and asked to make drawings of familiar objects. Total time of the experiment was typically about 30–40 min (fixed and adaptive procedures typically lasting about 5–10 min each).

D. Adaptive speech-based F0 discrimination of voice gender

An adaptive speech-based F0 discrimination procedure was used to assess voice gender perception as a function of voice F0 differences using an up-down staircase method (Levitt, 1971). Since several researchers have highlighted the susceptibility of adaptive threshold estimates to variability caused by attentional lapses or confusion at the beginning of trials (Baker and Rosen, 2001; Amitay *et al.*, 2006), subjects were first instructed how to respond in this task. They were told that they were going to hear two speech excerpts, one after the other, in which one item was always male and the other was always female, with no possibility that both were male or female. Their task was to listen to both speech items and choose by pressing the appropriate button on the keyboard which one was female. The female/male distinction was reinforced through analogy with the child's mother and father, and, a practice test was administered with ten trials. The majority of CI subjects successfully completed the practice test confirming that they understood the task; the others were all successful after repeating the practice test once more.

The stimuli were recordings of different utterances from the same speakers used in the identification procedure (no utterances were shared between procedures). In each trial, one speech item selected from 20 speakers of one sex was paired with another speech item selected from 20 speakers of the other sex. The position of the female voice in the stimulus pair was chosen at random. The stimulus set consisted of 400 different stimulus pairs, each of which had a unique ΔF0 between the female and male speech item. The largest separation of F0 in a stimulus pair was 142.8 Hz, and there was a region with negative ΔF0 for stimulus pairs in which the F0 of the female voice was lower than the F0 of the male voice (with the minimal value of $\Delta F0 = -23.6$ Hz).

The initial ΔF0 was set to 109 Hz, and the adaptive step was held constant at 10 Hz. The experiment stopped when either 10 reversal points had been achieved or 50 total trials had been run. In order to avoid possible pop-out memory effects of repeating stimulus pairs that were played recently, a minimal span of four trials was set between repeats of the same stimulus pair. In accordance with the relative scarcity of stimulus pairs on both extreme ends of the distribution of ΔF0, ΔF0 could not be increased or decreased beyond these edge points. To deal with cases in which subjects would become stuck at the edges, the adaptive procedure was modified; if the ΔF0 entered the region consisting of the five stimulus pairs with either the largest or the smallest F0 differences, then the adaptive procedure would randomly choose one of the stimulus pairs in this range.

Performance in the adaptive discrimination procedure was assessed with discrimination threshold estimates (DTEs). These were obtained for each subject by averaging the values of the last five reversal points of the adaptive procedure. Since the value of the adaptive parameter was limited by natural variation in the population of speakers, the subject might reach either a ceiling (indicating perfect performance) or floor level (indicating chance performance). In these cases, DTEs cannot be estimated more precisely than being smaller or larger than the edge boundaries. The lower boundary in this procedure was $\Delta F0 = -13.5$ Hz, whereas the upper boundary was $\Delta F0 = 133.1$ Hz.

Hearing controls were run using the same stimuli and procedures as implemented for the CI group, the only difference being the presentation of the sounds through Sennheiser HD 580 headphones [monaural presentation to the right ear at 65 dB SPL(*A*)].

E. Capturing procedures of stimulus-induced electrode output patterns from CI devices

In order to make sure that the CIs received direct stimulation from the experimental computer, without being contaminated by external background noise, a specially-constructed sound-insulated chamber was used (see Appendix). To maintain comparable loudness levels corresponding to 65 dB SPL(*A*) for all stimulus presentations to each subject, a two-step procedure was followed. The level of each stimulus was first calibrated by manual manipulation of signal levels in order to have a RadioShack model 332055 sound level meter reading of 65 dB SPL(*A*) in free-field, approximately 105 cm from the midline of a Harman Kardon 2.1 loudspeaker system. Then, using each subject's clinically-assigned CI processor in isolation from the subject, the master volume level on the computer playback software was adjusted by playing a standard 10-s multi-talker babble stimulus free-field via a Soundblaster Creative Audigy ZS sound card and Harman Kardon 2.1 loudspeaker system. The sound level meter was located at the Esprit 3G or Sprint microphone approximately 120 cm from the midline of the loudspeaker, and was used to measure the intensity level while adjusting the master/output level of the signal to produce a sound pressure level of 65 dB SPL(*A*). Once the proper setting was achieved, RF STATISTICS software was used to capture the CI output while playing each stimulus through the CI processor. Using the "Statistics" tab of the

capture software, it was possible to determine the average current level over the duration of each stimulus. The average of these current levels was taken, and the 10-s multi-talker babble stimulus was then played to the CI processor while the stimulus level was adjusted to produce an average current level identical to that obtained for the experimental stimuli. The CI processor was then returned to the subject, and this final setting was used to play the experimental stimuli to them. The relationship between the clinical units measured by the capturing system and actual physical (current) units (I in μA) was determined by the following equation:

$$I = F \times 10 \times 175^{\mathrm{cl}/255}, \quad (1)$$

where F is the calibration factor (equal to 1 for the devices studied here, Peter Seligman, Cochlear Ltd., personal communication) and cl is the value in clinical units (Drennan and Pfingst, 2006).

F. Analysis procedures for captured data

The 1 μs temporal resolution of the capturing apparatus was changed to 20 μs (corresponding to a sampling frequency of 50 kHz) using time-scale conversion, in order to decrease the memory burden for computational analyses. Each channel in each subject's device had minimum (threshold or T-level) and maximum (comfortable or C-level) current level values, which bracketed the device's stimulation levels, determined on the basis of the CI user's loudness percepts for pure tones during clinical fitting. These were standardized by conversion to percentage of dynamic range, using the formula

$$y = \frac{x - T}{C - T}, \quad (2)$$

where y is a converted value in percent of dynamic range, x is the stimulation current in units of current levels, T is the threshold level in units of current levels, and C is comfortable level in units of current levels for the stimulated electrode. The capturing procedure was not successful in 6 out of 41 subjects (denoted by asterisks in Table I); captured data therefore encompassed 35 subjects.

MSs were calculated via Fourier transformation of the autocorrelation of the stimulus output patterns (Singh and Theunissen, 2003). Bandpass filtering based on filter-banks using the same cut-off frequencies as in the electrodes in the CI device was first performed using the Nucleus Matlab Toolbox, a part of the NIC© proprietary software (Cochlear, 2002) generously provided by Cochlear AG. MSs were calculated by autocorrelating and then Fourier-transforming each envelope. This analysis included the frequency range between 75 and 225 Hz, which encompassed the F0 values of all stimuli [the upper bound of 225 Hz was one-quarter of the lowest stimulation rate (900 Hz), preventing aliasing effects in the data analysis]. The lower cut-off value of 75 Hz removed the effects of low-frequency modulation due to acoustic variation at syllabic/phonemic levels.

To measure the overall availability of F0 modulation information for a particular stimulus sound in each subject, the MSs of electrode outputs were summed over all channels that (1) were stimulated for at least 25% of the stimulus duration and (2) possessed distinct F0-related peaks in their acoustic MSs (peaks less than 40 Hz [a typical just-noticeable difference (JND) for CI temporal pitch for F0 values around 200 Hz (Zeng, 2002)] from the F0 value). To assess the availability of temporal cues for each stimulus item in each subject, the following algorithm was followed: (1) Find all local peaks in the MS that have energy at 0.8 or more of the maximum amplitude between 75 and 225 Hz and (2) find the closest peak to F0 with relatively greatest intensity. Using this procedure, each stimulus was assigned a CI MS F0-related peak value for each subject.

To assess the influence of the number of electrodes carrying F0-related modulation cues and their relative strength, cue-carrying electrodes were sorted according to the relative strength of their F0-related peaks ("modulation strength"). The analysis used the mean modulation strengths of the four strongest electrodes, and was performed both across response types (correct vs incorrect trials) and gender (male vs female speech items).

Individual DTEs fell into two broad groupings ($<$56 and $>$89 Hz), meaning that subjects with larger and smaller thresholds experienced a different number of trials and a different range of frequency differences. To assess the performance of their CI devices in an unbiased fashion, two different stimulus sets were selected based on the performance of subjects who scored above chance in the identification procedure. Stimulus set 1 (SS1) consisted of 15 stimulus pairs chosen from trials that were correctly responded to by 2 or more performing subjects and that were never incorrectly responded to by any of these subjects. Stimulus set 2 (SS2) contained 15 stimulus pairs chosen from trials responded to incorrectly by at least 2 performing subjects and that were never correctly responded to. The F0 separation between speech items in these trials was similar for the two sets and was limited to the range between 10 and 60 Hz ($\mathrm{mean}_{SS1} = 37.3 \pm 3.5$ Hz, $\mathrm{mean}_{SS2} = 31.7 \pm 3.6$, Mann–Whitney U-test, $Z = -0.975$, $n_{SS1} = 15$, $n_{SS2} = 15$, $p = 0.33$).

Finally, the quality of F0 cues to voice gender was quantified using the MS peak closest to F0, and the centroid position in the MS for each stimulus pair. Linear regressions were calculated using the F0 difference between stimuli as the independent variable, and the difference in CI MS peaks or MS-centroid positions as the dependent variable. The r^2 value for each subject was used as an estimator of the quality of information that each CI processor provided about each type of cue.

III. RESULTS

A. Voice gender identification

1. Behavioral data

Figure 2 shows mean voice gender identification results for the CI participants. Scores between 13 and 26 correct responses are consistent with chance performance based on the binomial distribution. Overall, 18 of the CI participants (44% of the CI subjects) could correctly identify the sex of the speaker at better-than-chance performance levels (mean value of $84.3 \pm 1.0\%$ correct) from an isolated stimulus;

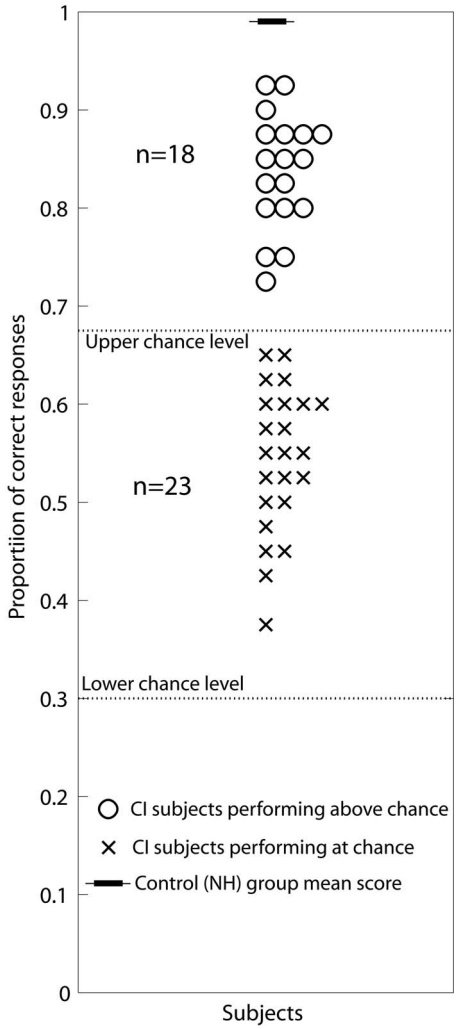


FIG. 2. Proportion of correct responses of 41 subjects with CIs in the fixed identification procedure. Dashed lines indicate the upper and lower boundaries for chance performance set by the binomial distribution. Performing CI subjects achieved a mean of $84.3 \pm 1.0\%$ correct responses; non-performing CI subjects had a mean of $54.4 \pm 2.0\%$ correct responses. The control group of NH subjects performed with a mean accuracy of $98.0 \pm 0.4\%$.

these subjects will subsequently be referred to as "performing" (P) subjects. The other 23 (56% of the study CI population), with a mean value of $54.4 \pm 2.0\%$ correct, were unable to identify the gender of isolated stimulus voices, and will subsequently be referred to as "non-performing" (NP) subjects.[2] All subjects in the control group performed with maximal or near-maximal performance (mean $98.0 \pm 0.4\%$ correct), confirming that the stimulus set was not of abnormal difficulty. Further analysis did not reveal any differences in CI subjects' performance between male and female speech items, independently of whether they could or could not correctly label them.

An analysis by items in subjects with above-chance performance is shown in Fig. 3. This reveals a V-shaped notch in the proportion of correct responses in relation to the speaker's F0, reflecting response uncertainty in the region of gender-ambiguous F0 values.[3]

2. Temporal F0-related modulation cues

The correlations between F0-related peaks measured from CI MSs and the corresponding F0 values measured from the stimulus waveforms were nominally significant in all 35 subjects that CI output signals were successfully captured from (15 performing and 20 non-performing), confirming that all CI devices provided robust F0-related temporal cues. The devices of performing and non-performing subjects did not significantly differ in the magnitudes of these correlations ($\text{mean}_P = 0.84$[SD=0.10], $\text{mean}_{NP} = 0.79$[SD=0.14], Mann–Whitney U-test, $Z = 0.98$, $n_P = 15$, $n_{NP} = 20$, $p = 0.325$). However, over all subjects, the correlation coefficients of correctly identified items were significantly higher than those of incorrectly identified items (means: 0.83 [SD=0.15] vs 0.54 [SD=0.48], Mann–Whitney U-test, $Z = 3.15$, $n = 35$, $p = 0.002$), indicating that the quality of temporal cues provided by the devices was significantly related to perceptual performance.

The 15 performing subjects had a large difference in correlation values between correct and incorrect items (means of 0.87 [SD=0.1] vs 0.21 [SD=0.56], Mann–Whitney U-test, $Z = 3.73$, $n = 15$, $p = 0.0002$). In contrast, there was no difference in the magnitude of the correlation coefficients between correct and incorrect items in the 20 non-performing subjects (means of 0.80 [SD=0.17] vs 0.78 [SD=0.18], Mann–Whitney U-test, $Z = 0.72$, $n = 20$, $p = 0.47$).

To combine the data from both types of voices into a single measure, an analysis of the distance of the F0-related peak in the CI MSs from the F0 "gender boundary" (the half-way point between the mean position of the male and female F0-related peaks of all speech items) and behavioral performance was carried out. All performing subjects had a positive correlation between these two measures, which was significantly different from 0 (Pearson correlation coefficients ranged from 0.33 and 0.83, $p < 0.05$ in all subjects; p-values were calculated using Fisher's r-to-z transform), suggesting that performance was related to how clear male or female temporal F0 cues were.

The acoustic and CI F0 distances of each item from the gender boundary were averaged across subjects, and pooled within performing and non-performing groups (Fig. 4). The F0 modulation information provided by the CI of the two groups is very similar, exhibiting a V-shaped relationship between acoustic and CI F0 distance from the gender boundary (this V-shape was previously observed in the analysis by items of the behavioral data from the performing subjects shown in Fig. 3). A regression of the form $y = B1 + B2 \times x$ between CI F0 distance and the proportion of subjects who correctly responded to each item was calculated for each subject group (where y is the proportion of subjects who responded correctly and x is the frequency distance of the F0-related peak in the CI MS from the gender boundary). The regression was significant and accounted for 24% of the observed gender identification variation in the performing CI subjects [$r^2 = 0.24$, $F(13) = 11.691$, $p = 0.002$], suggesting that they may use additional cues. The regression was not significant for the non-performing subjects [$r^2 = 0.05$, $F(18) = 1.913$, $p = 0.17$].

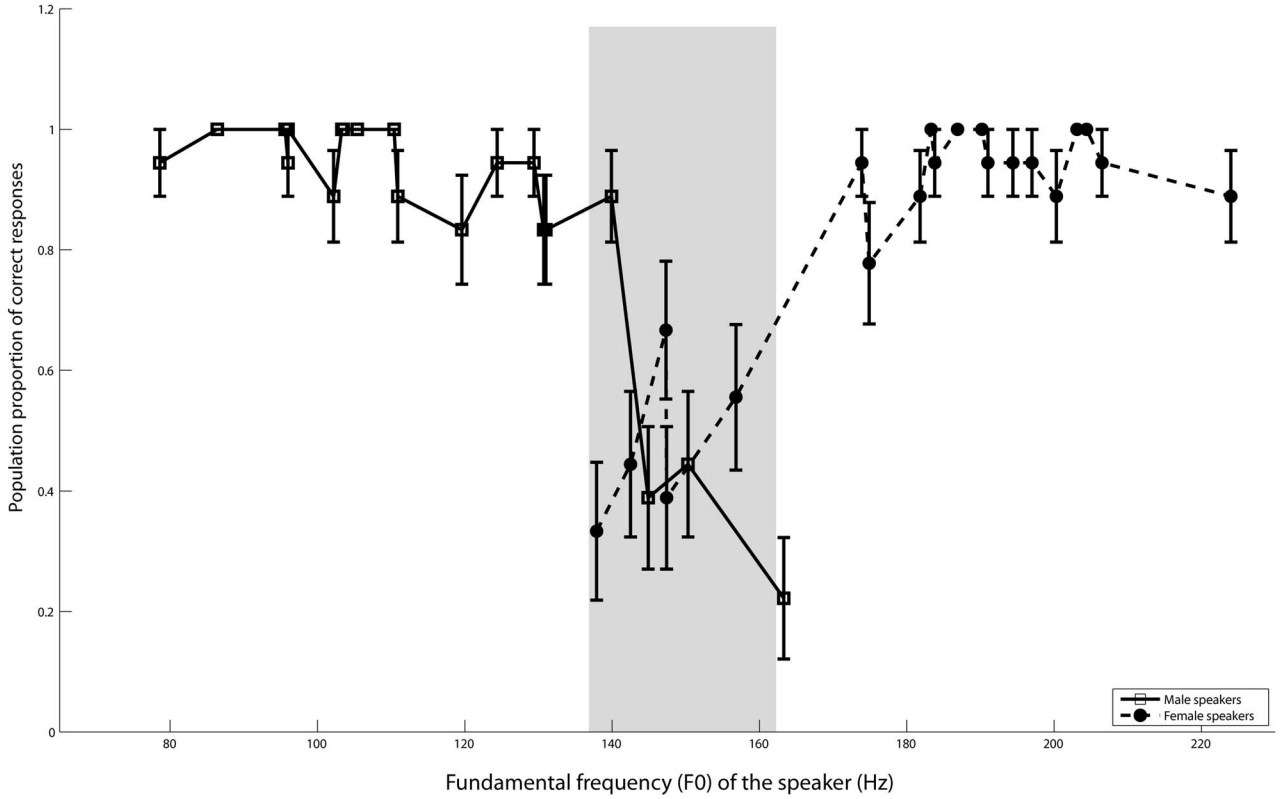


FIG. 3. Proportion of correct responses in performing subjects with CIs sorted by stimulus F0. The shaded area shows the region of overlapping F0 for male and female speech items. Error bars equal one standard error.

3. Number of channels carrying temporal modulation cues

In the gender identification task, the number of active electrodes was different between correctly-responded-to and incorrectly-responded-to trials in only 3 (two performing) of the 35 subjects from whom the capturing procedure was successful. Similar results were obtained with trials sorted according to the gender of the speaker (two subjects showed a significant difference in electrode numbers between male and female speech items). While a few subjects may show overall differences in the number of electrodes activated by male and female voices, it appears that this parameter did not play a general role in the identification task.

4. Magnitude of temporal modulation cues

The mean positions of electrodes with the four largest MS strength values were significantly different between female and male speech items in 4 (all performing) out of the 35 subjects. This suggests that in these four subjects there might be a relationship between the average position of the four electrodes having maximal MS strengths and voice F0. Two of these subjects (with relatively high performance) had a significant linear regression of MS strength on F0 with a negative slope for correctly-responded-to trials [subject CI24: $y=21.298-0.0042282X$, r^2(adjusted)$=0.18$, $F(30)=7.99$, $p=0.008$; subject CI31: $y=21.406-0.0050484X$, r^2(adjusted)$=0.17$, $F(33)=7.73$, $p=0.009$). Such an F0-MS strength relationship was not present for the other two subjects, who showed a greater variability of the mean electrode positions (ranges: 7.75–20 electrode units) compared to the former group (ranges: 19.5–21 electrode units). Still, on average, all four subjects had a systematic shift in the mean positions of the four electrodes containing the largest modulations between male and female items.

34 of the 35 subjects whose CI data were recorded showed a significant difference in the mean amount of modulation strength of the four most "active" electrodes between female and male speech items. Further analysis revealed a clear dependence of the strength of F0-related modulation and the value of F0. These equations were fitted with an exponential regression of the form $\mathrm{MS}_{\mathrm{strength}}=a\times e^{b\times F0}$. All 35 subjects showed a significant regression and there were no differences in the magnitude of the coefficients of determination between performing and non-performing subjects ($\mathrm{mean}_{\mathrm{P}}=0.54$[SD$=0.11$], $\mathrm{mean}_{\mathrm{NP}}=0.53$[SD$=0.13$], Mann–Whitney U-test, $Z=0.15$, $n_{\mathrm{P}}=15$, $n_{\mathrm{NP}}=20$, $p=0.88$). When this analysis was performed between correctly-responded-to and incorrectly-responded-to items, only one (non-performing) subject showed a difference in modulation strengths (means: 3.53×10^{-6} vs 6.11×10^{-6}, Mann–Whitney U-test, $Z=1.98$, n(correct)$=24$, n(incorrect)$=15$, $p<0.05$). In general, it seems that the subjects did not utilize these modulation strength differences while performing gender identification, even though speech items with lower F0s produced stronger F0-related modulations in all CI devices.

5. Place cues

The mean male-female spatial distance for all participating subjects was 0.28 ± 0.03 electrode units and is significantly different from 0 (Wilcoxon matched-pair sign rank

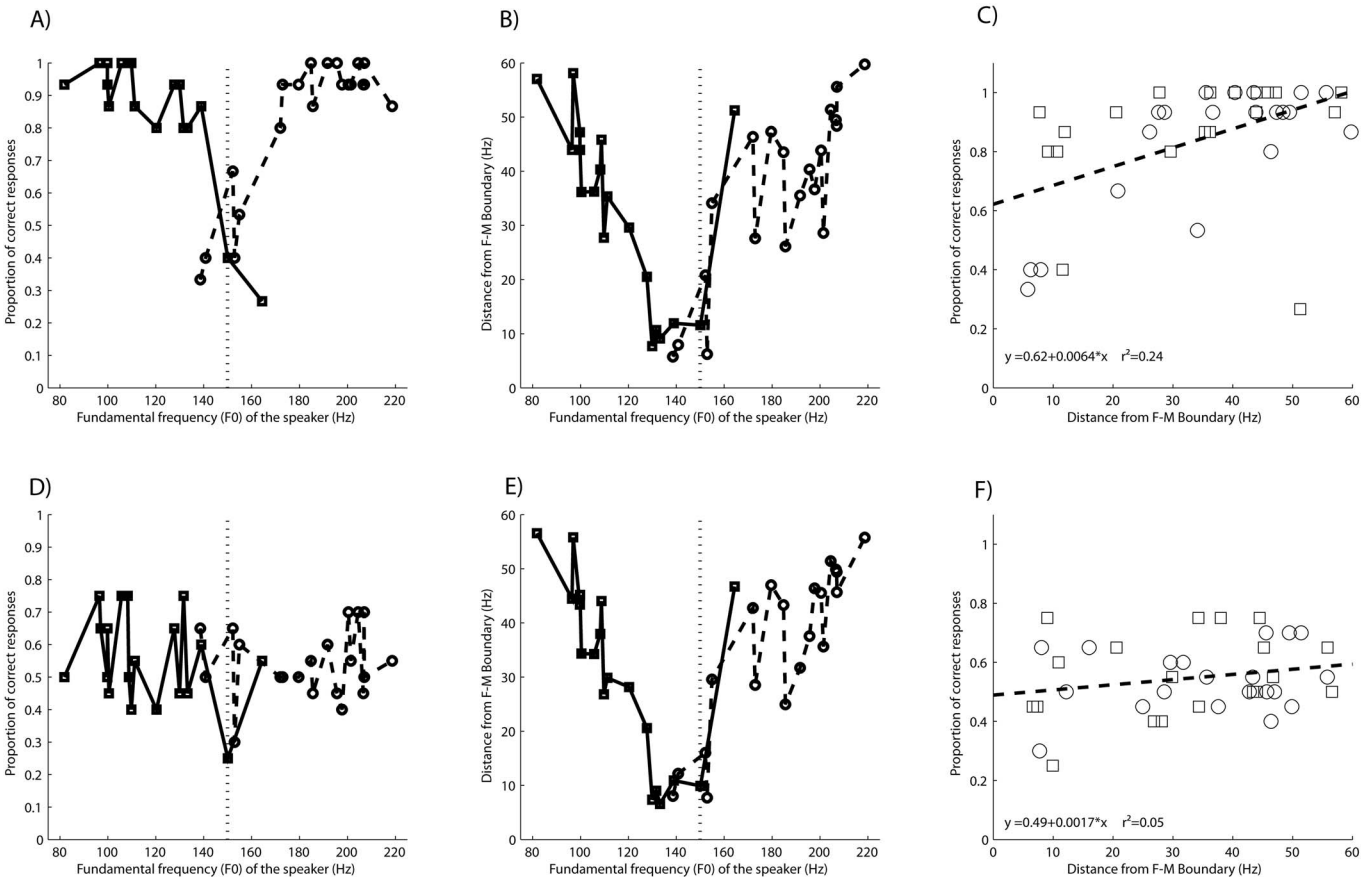


FIG. 4. The availability of F0 temporal cues provided by the CI as a function of F0, and its relation to behavioral performance. Upper panels (A)–(C) show data for the performing group of subjects, while lower panels (D)–(F) show results for the non-performing subjects. In all panels, circles denote female speech items, and squares denote male speech items. Panels (A) and (D) show the relationship between the proportions of subjects who responded correctly and the F0 of the speech item. Panels (B) and (E) show average F0 distances from the gender boundary as recorded from the CI output patterns for each speech item. Panels (C) and (F) show the relationship between the proportion of subjects who correctly identified each speech item and the corresponding F0 distance from the gender boundary as recorded from the CI output patterns. The dashed line represents the linear regression. Coefficients of linear regression and their corresponding coefficients of determination are given on the bottom left of the panel.

test, $T^+=594$, $Z=-4.57$, $n=35$, $p<0.001$). The mean spatial difference in the devices of performing group subjects was statistically the same as the difference in the devices of non-performing subjects (Mann–Whitney U-test, $Z=-0.1167$, $n_P=15$, $n_{NP}=20$, $p=0.91$). Although place cues were available to all subjects, they were of small magnitude. An additional analysis using only the ten most extreme F0 values in each gender group still showed no difference in the magnitude of place cues between the devices of performing and non-performing subjects (Mann–Whitney U-test, $Z=-0.38$, $n_P=15$, $n_{NP}=20$, $p=0.70$). Analyzing place data by response type (correctly-responded-to items and incorrectly-responded-to items) again revealed no differences in the devices of either group of CI subjects (performing group, all items: U-test, $Z=0.55$, $n_C=15$, $p_C=0.59$; non-performing group, all items: U-test, $Z=0.47$, $n_I=20$, $p_I=0.64$). Place cue distances from the gender boundary were not associated with acoustic F0 frequency distances in any of the subjects. While subtle information about center-of-mass place cues is available in the devices of all subjects, it seems to be of little utility for voice gender identification.

Place cues may provide CI subjects with additional information in the form of consistent covariation with temporal cues, which could be masked by the stronger salience of temporal information. However, a joint analysis assessing the relationship between temporal (F0-related peak in MS) and spectral (CM positions of the stimulus output pattern) distances from the gender boundary for all subjects revealed no significant relationships between place and temporal cues (analysis not shown).

B. Adaptive speech-based F0 discrimination of voice gender

1. Behavioral data

The control group performed perfectly up to the limits of the experimental stimulus pairs (12 subjects did not show reversals, with the remaining 3 subjects showing two reversals only at $\Delta F0<0$). This demonstrates that the task works well for hearing children. The picture is markedly different for the CI users. Figure 5 shows the DTEs for each subject. Two groups of subjects are readily apparent: (i) 23 CI children produced small ("good") DTEs with values of 55.8 Hz or less, and (ii) the other 18 CI subjects had substantially higher ("worse") DTEs (89.8 Hz and higher). Threshold in this testing scheme corresponds to a performance of 70.7% correct.

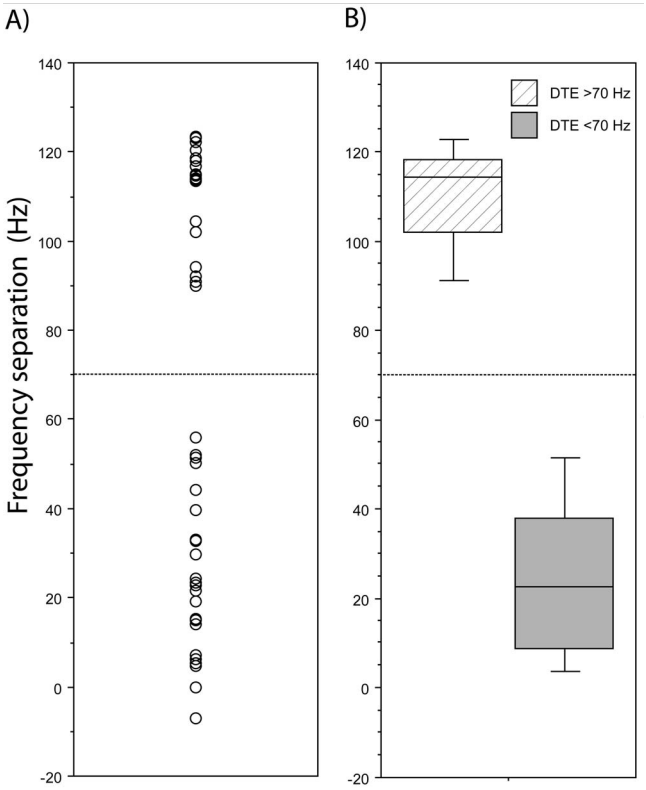


FIG. 5. Scores in the adaptive speech-based F0 discrimination procedure: (A) adaptive DTEs for each CI subject; (B) box plot of the DTE for both performance groups. In-box horizontal lines are the median values. Plotting bar boundaries are between the 25%th and 75%th percentiles; error bars go between the 10%th and 90%th percentiles.

2. Temporal cues

Performing and non-performing subjects[4] had similar r^2 values for both measures of F0 information in both stimulus sets [Mann–Whitney U tests, stimulus set 1: r^2(MS-centroid)$=0.59\pm0.03$ vs 0.54 ± 0.04, $Z=-1.20$, $n_P=19$, $n_{NP}=14$, $p=0.23$; r^2(F0-peak)$=0.58\pm0.04$ vs 0.54 ± 0.07, $Z=-0.22$, $n_P=19$, $n_{NP}=14$, $p=0.82$; stimulus set 2: r^2(MS-centroid)$=0.31\pm0.04$ vs 0.35 ± 0.04, $Z=-0.49$, $n_P=19$, $n_{NP}=14$, $p=0.62$; r^2(F0-peak)$=0.22\pm0.03$ vs 0.21 ± 0.05, $Z=-0.26$, $n_P=19$, $n_{NP}=14$, $p=0.8$]. This empirically confirms that the CI devices of performing and non-performing subjects provided similar F0 information.

Figure 6 shows differences between the stimulus sets in the quality of the information transmitted by the CI. Items from stimulus set 1 provide significantly better temporal information than items from stimulus set 2 in all subjects (Wilcoxon matched-pair sign rank test for F0-peak and MS-centroid information: $T^+=538$, $Z=-4.60$, $n=33$, $p<0.0001$). These differences exist in both performance groups for both temporal measures (Wilcoxon matched-pair sign rank test, F0-peak information; performing group, $T^+=189$, $Z=-3.78$, $n=19$, $p=0.0002$; non-performing: $T^+=96$, $Z=-2.73$, $n=14$, $p=0.0063$; MS-centroid information: performing group, $T^+=186$, $Z=-3.66$, $n=19$, $p=0.0003$; non-performing: $T^+=96$, $Z=-2.73$, $n=14$, $p=0.0063$).

Another way to look at the quality of temporal information provided by the CI is the degree of consistency between the CI MS spectra peaks and MS-centroid positions, as measured by their correlation. Over all subjects, the correlation coefficient was nominally larger in stimulus set 1 compared to stimulus set 2 [0.57 ($n=33$, $p<0.001$, Fisher r-to-z) vs 0.39 ($n=33$, $p=0.02$, Fisher r-to-z)], but the difference in the magnitude of these two correlation coefficients was not significant ($z=0.91$, $p=0.35$). Similar results were found when this analysis was repeated separately for each performance group (analysis not shown).

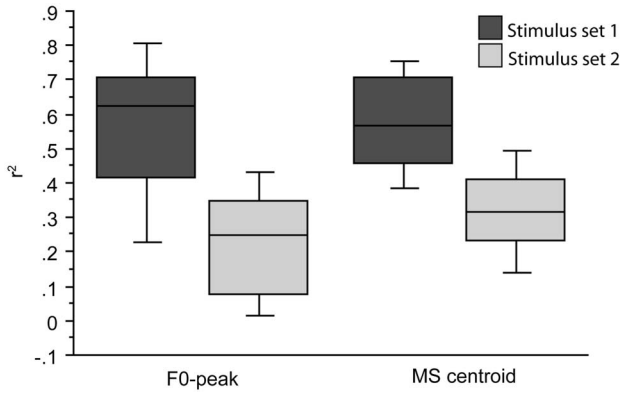


FIG. 6. Differences in r^2 between stimulus sets for measures of temporal information provided by the CI device (F0-related peak, left; and MS-centroid, right). In-box horizontal lines are the median values. Plotting bar boundaries are between the 25%th and 75%th percentiles; error bars go between the 10%th and 90%th percentiles.

In order to assess whether either of the temporal measures provided better cues than the other, a comparison of the r^2 values for items from stimulus set 1 was examined, but this revealed no advantage over all subjects, or for the performing or non-performing groups. The same analysis for stimulus set 2 revealed that MS-centroids were represented with a higher degree of fidelity compared to F0-peak information [Wilcoxon signed rank test, performing group: r^2(MS-centroid)$=0.31\pm0.04$ vs r^2(F0-peak)$=0.22\pm0.03$, $T^+=41$, $Z=-2.173$, $n=19$, $p=0.03$; non-performing group: r^2(MS-centroid)$=0.35\pm0.04$ vs r^2(F0-peak)$=0.21\pm0.05$, $T^+=18$, $Z=-2.166$, $n=14$, $p=0.03$; all subjects: r^2(MS-centroid)$=0.33\pm0.03$ vs r^2(F0-peak)$=0.22\pm0.03$, $T^+=112$, $Z=-3.011$, $n=33$, $p=0.003$). The fact that MS-centroid information is better-represented in stimulus items that CI subjects discriminate more poorly suggests that CI users may not be able to extract information from this cue very effectively.

Adaptive discrimination performance could be limited by the availability of relevant temporal information. However, an examination of regressions between the r^2 values of MS-centroid and F0-peak temporal information and adaptive DTEs yielded no significant relationships.

3. Place cues

The analysis of place cues in the form of the stimulus pair differences between center-of-mass position of the stimulus output patterns and the corresponding F0 revealed that only two subjects had significant correlations for correctly-responded items (performing subject CI30, correlation coefficient=0.40, $p=0.02$, and $n=34$; non-performing subject CI25, correlation coefficient=0.48, $p=0.05$, and n

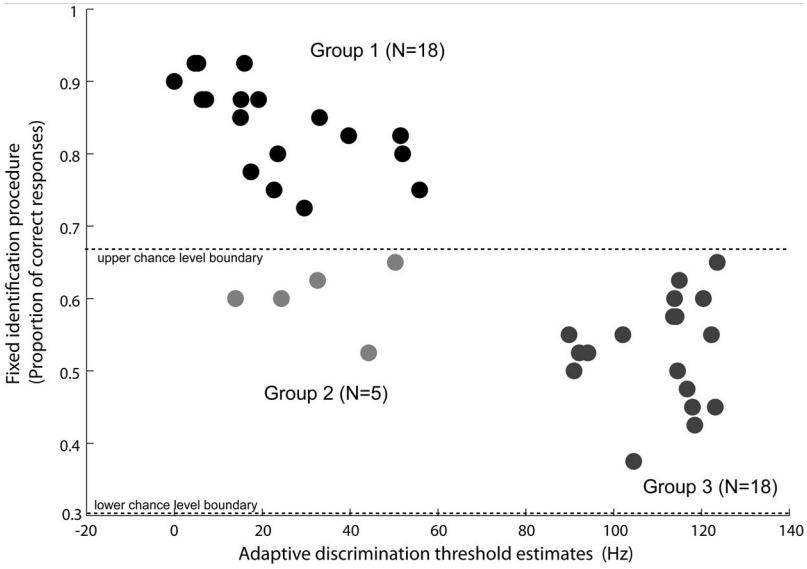


FIG. 7. Two-dimensional plot of the joint results of the fixed identification and adaptive speech-based F0 discrimination procedures. Each data point represents one subject.

=17). As in the identification task, temporal cues in the form of amplitude modulation of pulse trains appear to be a major source of useful information about speaker gender that CI devices are transmitting.

C. Joint results of the fixed identification and adaptive discrimination procedures

Figure 7 illustrates the joint results of the two procedures used in this study. Over all subjects, there is strong correlation between the performances obtained from the two different assessments (correlation=−0.822, $n=41$, $p<0.0001$, Fisher r-to-z). The data indicate three natural groupings of CI subjects based on their task performance. Group 1 (upper left) consists of 18 subjects with above-chance performance in voice gender identification and adaptive DTEs below 70 Hz. These subjects were able to identify gender at above-chance levels and performed at 70.7% or better in the adaptive discrimination task. Group 2 consists of five subjects (lower left) who were unable to correctly identify voice gender in spite of having speech-based F0 DTEs between 15 and 60 Hz, overlapping completely with many of the subjects in group 1. Group 3 consists of 18 subjects who were not able to correctly identify voice gender, and showed poor adaptive speech-based F0 DTEs as well.

One possibility for the differential performance in the two experiments could be that the subjects from group 2 learned to identify voice gender in the course of the identification procedure (which always preceded the discrimination experiment, and had feedback to maintain the subject's interest). However, none of these subjects showed significant differences in performance for the initial and final blocks of speech items in the identification procedure when assessed with Mann–Whitney tests, suggesting that learning did not occur.

IV. DISCUSSION

This study evaluated how children with CIs performed in gender identification tasks using naturalistic stimuli, and empirically assessed the quality of information that their implant devices were providing to them. In the fixed identification procedure, 18 out of 41 subjects (44%) could identify gender from 2-s speech items taken from a set representing 20 male and 20 female speakers, whereas the other 23 (56%) of the subjects were unable to perform this task. The results demonstrate that individual variation in the use of CI information is not only seen in complex cognitive tasks such as speech perception (Blamey *et al.*, 2001), but is also observed in more basic perceptual tasks such as identification of voice gender.

There are several exogenous and endogenous factors that might contribute to this variation. Major exogenous factors include technical and clinical aspects of cochlear implantation (the device and the surgery), whereas the major endogenous factors include the fundamental properties of the auditory system together with the developmental trajectory of deafness and CI use. A major but little-addressed issue examined by the present study concerns the quality of information that CI devices provide to their users.

Patterns of stimulation delivered by each subject's CI device were examined, and no general difference was found in the quality of information provided by the devices of subjects who perform gender identification well and those who do not; performance differences appear to result from the ability of CI users to make use of these cues. This variation in the ability to use information may result from differential effects of electrode position and insertion depth, and interactions with cochlear biophysics (which are likely to be different in different subjects), resulting in differences in the number, location, and quality of stimulation of auditory nerve fibers. All the CI users in the present study underwent full-depth electrode insertion, and it was not possible to compare the relatively small electrode insertion depth differences among them with voice gender identification performance.

However, data were collected on exogenous variables such as age of the CI surgery, age of the onset of deafness, duration of CI use, and duration of deafness; an evaluation of the role of these variables will appear in a separate publication.

For the group of CI wearers who were proficient at gender identification, the quality of temporal information was significantly better for the items that they responded to correctly when compared to incorrect items. Such a contrast was not found in the group of subjects who performed poorly in gender identification, suggesting that the performing group of subjects could use the temporal information provided by their devices. Place cues were available but of small magnitude, with no evidence for their use. A greater use of temporal cues over spectral cues has also been suggested by previous researchers. For example, Fu *et al.* (2004) compared gender discrimination abilities using vowels, and found up to 30% better performance when temporal cues were available. Fu *et al.* (2005) compared gender identification abilities with vowels spoken by talkers coming from two sets, with an F0 separation between female and male speakers of 100 and 10 Hz. CI users performed well with the first talker set (F0 separation of 100 Hz), but their performance decreased sharply for the set of speakers with small F0 separation.

The issue remains whether the non-utilization of place information in the experiments presented here is due to difficulties in detection, or masking by the stronger presence of temporal cues. A preference for temporal over place cues for fundamental frequency was confirmed by Laneau and Wouters (2004b), who found an increase in just-noticeable differences for F0 discrimination in four CI users from just above 1 octave when place cues were presented alone, to 6%–60% of an octave when temporal information was added.

The fact that F0 information only accounts for approximately one-quarter of the variation in voice gender identification [Fig. 4(C)] suggests the use of additional gender-related cues. One such cue was present in the form of variation in electrode modulation strength. All subjects' devices exhibited a significant relationship between the modulation strength of the four most active electrodes and voice F0, but subjects appeared insensitive to it. It is unclear whether modulation strength cues are not useful because they are perceptually difficult to detect, because CI users are unable to interpret them, or because users are not paying attention to them; it is equally unclear whether CI users could learn to make use of them. Further research with CI subjects and NH subjects listening to CI simulations is necessary to better understand and exploit any perceptual relevance of electrode modulation strengths and their spatial patterning for sound processing schemes.

The CI speech coding algorithm implemented in the devices studied here does not appear to provide equally good temporal representations of fundamental frequencies for all voices. Stimulus pairs that were more problematic for gender identification also had poorer temporal representations of fundamental frequency. Since CI output patterns are the result of a complex interplay between clinical parameter settings of the devices and spectro-temporal features of the incoming sounds, this suggests the need for a better optimization strategy for the speech coding algorithm in order to facilitate gender identification for a wider variety of voices.[5]

Finally, this study identified five CI users who could not identify the gender of singly-presented voices but could do so at better-than-chance levels when two voices were sequentially presented. This dissociation of abilities for voice gender identification and adaptive speech-based F0 discrimination is striking, as these users had periods of CI experience (an average of 4.04 ± 0.96 years of use) equivalent to the 18 CI users who could successfully perform both tasks (an average of 4.02 ± 1.06 years of use).

The two paradigms differ in terms of how subjects might strategically perform them. Auditory events appear to produce a transient auditory sensory memory trace lasting up to 10–20 s (Cowan, 1984; Semal and Demany, 1991; Krumhansl and Iverson, 1992; Clement *et al.*, 1999; Caclin *et al.*, 2006). In a single-interval gender identification task, gender-relevant acoustic information in a stimulus item must be compared with long-term memory representations of gender-related acoustic features. Two items presented one after the other within ~5 s could be compared without reference to long-term representations. Winkler *et al.* (2002) and Winkler and Cowan (2005) assessed ways in which short-term auditory ("surface") information gets encoded into longer-term memory, suggesting that the auditory context within which the sounds are to be remembered plays an important role. Perhaps some general types of auditory experiences at an early age are necessary to set up relations between sounds and contexts that facilitate the formation of long-term categorical auditory memories. This issue has important implications for the ease with which CI users may be able to use different kinds of auditory information, and needs to be examined more rigorously in future research.

V. CONCLUSION

Forty-one juvenile cochlear implantees were exposed to naturalistic speech samples from a variety of speakers, and asked to perform two voice gender perception tasks; the stimulus output patterns of their CIs were also documented for each speech sample. CI electrical output features related to voice fundamental frequency (F0) and spectral envelope were evaluated in relation to behavioral performance. Temporal and place cues were equally available in all CI devices, but only about half of the subjects were able to label gender correctly. Subjects who could identify voice gender appeared to utilize temporal cues but showed no evidence of using place cues. At least one other robust F0-related cue was present in the output of all CI devices that participants appeared unable to make use of. A subgroup of participants could discriminate voice gender when two contrasting voices were presented in succession, but were unable to identify gender when voices were singly presented, suggesting that it may be fruitful to more carefully examine the characteristics of long-term auditory category formation in CI user populations.

ACKNOWLEDGMENTS

This work was supported by the Croatian Ministry of Science, Education and Sports Grant (Grant No. 207-0000000-2293), the Central European Initiative Science and Technology Network Research Fellowship to D.K., NSERC Grant No. 298612 and CFI Grant No. 9908 to E.B., and SISSA. The authors thank Ernst von Wallenberg and Cochlear AG, Basel for providing hardware, Andrew Vandali (CRC HEAR, Melbourne) and Peter Seligman (Cochlear Ltd., Melbourne) for software and consultation, Davor Petrinović (Faculty of Electrical Engineering and Computing, Zagreb) and Vladimir Kozina (AKO Electrical Engineering, Zagreb) for technical support and the staff of SUVAG Polyclinic for logistic support, and Colette McKay and two anonymous reviewers for helpful comments.

APPENDIX: COCHLEAR IMPLANT MICROPHONE ISOLATION

For subjects using Sprint processors, an Intra-Operative direct cable (manufactured by Cochlear Ltd., Sydney, Australia and provided by Cochlear AG, Basel, Switzerland) that shuts off the external microphone was used during these experiments. However, most of the subjects used an Esprit 3G CI processor (typically programed with the mixing mode enabled). Direct analog input to the processor via a cable connection would not normally switch off the external processor's microphone, resulting in the analog input being mixed with environmental sounds. This was undesirable during the sound measurements and experimental stimulus presentations, and it was equally undesirable to change any parameter settings on a subject's processor. Therefore, an acoustic isolation chamber for the implant ear-piece was constructed. This consisted of a small thermos double-walled container filled with cement aggregated with small iron balls (1–2 mm in diameter) in order to obtain high density for reducing sound transmission. The Esprit 3G processor was placed within the insulated box, which was then sealed. A 30-cm long coil cable was connected to the processor. The insulated box was suspended in the vicinity of and behind the subject's head, making sure that it did not distract the subject or interfere with their comfort.

The attenuation of the insulated box was measured with an internal microphone (ECM 335-62M frequency range 20–20 000 Hz) and an external source of white noise [95 dB SPL (*C*-weighting, slow reading)] using three different frequency ranges: (1) octave-wide filtered white noise (60 dB/octave roll-off) centered at 1000 Hz (27 dB attenuation), (2) octave-wide filtered white noise (60 dB/octave roll-off) centered at 2000 Hz (25 dB attenuation), and (3) stimulation by a loudspeaker with a frequency range 100–20 000 Hz (30 dB attenuation). In addition, a generic CI device (Esprit 3G, 15 active electrodes, frequency table 7, frequency range 120–8658 Hz, and stimulation rate: 900 Hz pps/channel) sealed inside the box required augmentation of the sound by 36 dB to produce similar average pulse output activity.

[1]The F0 values of the speech items used here range from 80 to 220 Hz; CI maps typically have this entire frequency range assigned to the most apical active electrode. How is it possible to use any F0 place cue information when it all falls within one channel? F0 differences are also correlated with differences in the distribution of spectral envelopes between male and female speakers (Ives *et al.*, 2005; Smith and Patterson, 2005; Smith *et al.*, 2005, 2007). In the course of auditory experience with the device, CI users might build up a representation of the population distribution of the center-of-mass of stimulation over the electrode array for male and female voices, which could allow them to discern stimulation patterns whose center-of-mass differs by fractions of an electrode position.

[2]If the male and female voices with an overlapping F0 are excluded from the analysis, the mean performance increased significantly to $94.1 \pm 2.0\%$ in the performing group of CI subjects (Mann–Whitney U-test, $n=18$, $Z=-4.02$, and $p<0.000\,01$), but was unchanged in the non-performing group of CI subjects (mean performance $56.0 \pm 2.0\%$, Mann–Whitney U-test, $n=23$, $Z=-0.26$, and $p=0.79$).

[3]One male speech item was consistently ambiguous in the control group, and this item was excluded from all further analyses.

[4]CI output patterns captured from stimulus sets 1 and 2 were unusable in two non-performing subjects because of equipment failure.

[5]Cleary and Pisoni (2002) found that CI users performed poorly in voice identification when the linguistic content of paired utterances differed. This does not necessarily conflict with the present findings. In Cleary and Pisoni, 2002, the subjects performed a "same-different" task for identifying the speaker (using 3 different female voices) in the face of variation in the content of utterances, in contrast to the voice gender identification task using 20 male and 20 female voices studied here. Short-term comparison may have facilitated the performance of the subjects in the present study, but not those in Cleary and Pisoni, 2002, because gender identification can be successfully based on a few relatively static vocal cues (even with a large number of voices). In contrast, useful cues for speaker identification include more dynamic voice features, and subjects need to use different combinations of static and dynamic features to tell the difference between different types of voices. Short-term comparisons would help with the former kind of task, but may hinder overall performance in the latter kind of task.

Amitay, S., Irwin, A., Hawkey, D. J., Cowan, J. A., and Moore, D. R. (**2006**). "A comparison of adaptive procedures for rapid and reliable threshold assessment and training in naive listeners," J. Acoust. Soc. Am. **119**, 1616–1625.

Bachorowski, J. A., and Owren, M. J. (**1999**). "Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech," J. Acoust. Soc. Am. **106**, 1054–1063.

Baker, R. J., and Rosen, S. (**2001**). "Evaluation of maximum-likelihood threshold estimation with tone-in-noise masking," Br. J. Audiol. **35**, 43–52.

Blamey, P. J., Sarant, J. Z., Paatsch, L. E., Barry, J. G., Bow, C. P., Wales, R. J., Wright, M., Psarros, C., Rattigan, K., and Tooher, R. (**2001**). "Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing," J. Speech Lang. Hear. Res. **44**, 264–285.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," Proc. Inst. Phonetic Sci. **17**, 97–110.

Boersma, P., and Weenink, D. (**2006**). *Praat: Doing Phonetics by Computer (Version 4.4.32)* [Computer program]. Retrieved October 1, 2006 from http://www.praat.org/.

Burns, E. M., and Viemeister, N. F. (**1976**). "Non-spectral pitch," J. Acoust. Soc. Am. **60**, 863–869.

Burns, E. M., and Viemeister, N. F. (**1981**). "Played-again SAM: Further observations on the pitch of amplitude-modulated noise," J. Acoust. Soc. Am. **70**, 1655–1660.

Caclin, A., Brattico, E., Tervaniemi, M., Naatanen, R., Morlet, D., Giard, M. H., and McAdams, S. (**2006**). "Separate neural processing of timbre dimensions in auditory sensory memory," J. Cogn. Neurosci. **18**, 1959–1972.

Carlyon, R. P., and Deeks, J. M. (**2002**). "Limitations on rate discrimination," J. Acoust. Soc. Am. **112**, 1009–1025.

Chang, Y. P., and Fu, Q. J. (**2006**). "Effects of talker variability on vowel recognition in cochlear implants," J. Speech Lang. Hear. Res. **49**, 1331–1341.

Childers, D. G., and Wu, K. (**1991**). "Gender recognition from speech. Part II: Fine analysis," J. Acoust. Soc. Am. **90**, 1841–1856.

Cleary, M., and Pisoni, D. B. (**2002**). "Talker discrimination by prelingually

deaf children with cochlear implants: Preliminary results," Ann. Otol. Rhinol. Laryngol. Suppl. **189**, 113–118.
Cleary, M., Pisoni, D. B., and Kirk, K. I. (**2005**). "Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants," J. Speech Lang. Hear. Res. **48**, 204–223.
Clement, S., Demany, L., and Semal, C. (**1999**). "Memory for pitch versus memory for loudness," J. Acoust. Soc. Am. **106**, 2805–2811.
Cochlear (**2002**). "Nucleus implant communicator (NIC) system overview," Cochlear Ltd.
Cohen, L. T., Busby, P. A., Whitford, L. A., and Clark, G. M. (**1996**). "Cochlear implant place psychophysics 1. Pitch estimation with deeply inserted electrodes," Audiol. Neuro-Otol. **1**, 265–277.
Cowan, N. (**1984**). "On short and long auditory stores," Psychol. Bull. **96**, 341–370.
Drennan, W. R., and Pfingst, B. E. (**2006**). "Current-level discrimination in the context of interleaved, multichannel stimulation in cochlear implants: Effects of number of stimulated electrodes, pulse rate, and electrode separation," J. Assoc. Res. Otolaryngol. **7**, 308–316.
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., Chinchilla, S., Nogaki, G., and Galvin, J. J. (**2005**). "Voice gender identification by cochlear implant users: The role of spectral and temporal resolution," J. Acoust. Soc. Am. **118**, 1711–1718.
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.
Holt, R. F., and Kirk, K. I. (**2005**). "Speech and language development in cognitively delayed children with cochlear implants," Ear Hear. **26**, 132–148.
Ives, D. T., Smith, D. R. R., and Patterson, R. D. (**2005**). "Discrimination of speaker size from syllable phrases," J. Acoust. Soc. Am. **118**, 3816–3822.
Krumhansl, C. L., and Iverson, P. (**1992**). "Perceptual interactions between musical pitch and timbre," J. Exp. Psychol. Hum. Percept. Perform. **18**, 739–751.
Laneau, J., and Wouters, J. (**2004a**. "Multichannel place pitch sensitivity in cochlear implant recipients," J. Assoc. Res. Otolaryngol. **5**, 285–294.
Laneau, J., and Wouters, J. (**2004b**). "Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees," J. Acoust. Soc. Am. **116**, 3606–3619.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
McDermott, H. J., and McKay, C. M. (**1994**). "Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea," J. Acoust. Soc. Am. **96**, 155–162.
McKay, C. M. (**2005**). "Spectral processing in cochlear implants," in *Auditory Spectral Processing*, edited by M. S. Malmierca and D. R. F. Irvine (Elsevier Academic, San Diego, CA), pp. 473–509.
McKay, C. M., and McDermott, H. J. (**2000**). "Place and temporal cues in pitch perception: Are they truly independent?," ARLO **1**, 25–30.
McKay, C. M., McDermott, H. J., and Clark, G. M. (**1994**). "Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees," J. Acoust. Soc. Am. **96**, 2664–2673.
McKay, C. M., McDermott, H. J., and Clark, G. M. (**1995**). "Pitch matching of amplitude-modulated current pulse trains by cochlear implantees: The effect of modulation depth," J. Acoust. Soc. Am. **97**, 1777–1785.
Moulines, E., and Laroche, J. (**1995**). "Non-parametric techniques for pitch-scale and time-scale modification of speech," Speech Commun. **16**, 175–205.
Nicholas, J. G., and Geers, A. E. (**2007**). "Will they catch up? The role of age at cochlear implantation in the spoken language development of children with severe to profound hearing loss," J. Speech Lang. Hear. Res. **50**, 1048–1062.
Owren, M. J., Berkowitz, M., and Bachorowski, J. A. (**2007**). "Listeners judge talker sex more efficiently from male than from female vowels," Percept. Psychophys. **69**, 930–941.
Semal, C., and Demany, L. (**1991**). "Dissociation of pitch from timbre in auditory short-term memory," J. Acoust. Soc. Am. **89**, 2404–2410.
Singh, N. C., and Theunissen, F. E. (**2003**). "Modulation spectra of natural sounds and ethological theories of auditory processing," J. Acoust. Soc. Am. **114**, 3394–3411.
Smith, D. R. R., and Patterson, R. D. (**2005**). "The interaction of glottal-pulse rate and vocal-tract length in judgments of speaker size, sex, and age," J. Acoust. Soc. Am. **118**, 3177–3186.
Smith, D. R. R., Patterson, R. D., Turner, R., Kawahara, H., and Irino, T. (**2005**). "The processing and perception of size information in speech sounds," J. Acoust. Soc. Am. **117**, 305–318.
Smith, D. R. R., Walters, T. C., and Patterson, R. D. (**2007**). "Discrimination of speaker sex and size when glottal-pulse rate and vocal-tract length are controlled," J. Acoust. Soc. Am. **122**, 3628–3639.
Spahr, A. J., and Dorman, M. F. (**2004**). "Performance of subjects fit with the Advanced Bionics CII and Nucleus 3G cochlear implant devices," Arch. Otolaryngol. Head Neck Surg. **130**, 624–628.
Stacey, P. C., Fortnum, H. A., Barton, G. R., and Summerfield, A. Q. (**2006**). "Hearing-impaired children in the United Kingdom, I: Auditory performance, communication skills, educational achievements, quality of life, and cochlear implantation," Ear Hear. **27**, 161–186.
Svirsky, M. A., Robbins, A. M., Kirk, K. I., Pisoni, D. B., and Miyamoto, R. T. (**2000**). "Language development in profoundly deaf children with cochlear implants," Psychol. Sci. **11**, 153–158.
Svirsky, M. A., Teoh, S. W., and Neuburger, H. (**2004**). "Development of language and speech perception in congenitally, profoundly deaf children as a function of age at cochlear implantation," Audiol. Neuro-Otol. **9**, 224–233.
Waltzman, S. B., Scalchunes, V., and Cohen, N. L. (**2000**). "Performance of multiply handicapped children using cochlear implants," Am. J. Otol. **21**, 329–335.
Winkler, I., and Cowan, N. (**2005**). "From sensory to long-term memory—Evidence from auditory memory reactivation studies," J. Exp. Psychol. **52**, 3–20.
Winkler, I., Korzyukov, O., Gumenyuk, V., Cowan, N., Linkenkaer-Hansen, K., Ilmoniemi, R. J., Alho, K., and Naatanen, R. (**2002**). "Temporary and longer term retention of acoustic information," Psychophysiology **39**, 530–534.
Wu, K., and Childers, D. G. (**1991**). "Gender recognition from speech. Part I: Coarse analysis," J. Acoust. Soc. Am. **90**, 1828–1840.
Zeng, F. G. (**2002**). "Temporal pitch in electric hearing," Hear. Res. **174**, 101–106.

Perceptually relevant parameters for virtual listening simulation of small room acoustics

Pavel Zahorik[a)]
Department of Psychological and Brain Sciences, University of Louisville, Louisville, Kentucky 40292

(Received 12 August 2008; revised 22 May 2009; accepted 11 June 2009)

Various physical aspects of room-acoustic simulation techniques have been extensively studied and refined, yet the perceptual attributes of the simulations have received relatively little attention. Here a method of evaluating the perceptual similarity between rooms is described and tested using 15 small-room simulations based on binaural room impulse responses (BRIRs) either measured from a real room or estimated using simple geometrical acoustic modeling techniques. Room size and surface absorption properties were varied, along with aspects of the virtual simulation including the use of individualized head-related transfer function (HRTF) measurements for spatial rendering. Although differences between BRIRs were evident in a variety of physical parameters, a multidimensional scaling analysis revealed that when at-the-ear signal levels were held constant, the rooms differed along just two perceptual dimensions: one related to reverberation time (T_{60}) and one related to interaural coherence (IACC). Modeled rooms were found to differ from measured rooms in this perceptual space, but the differences were relatively small and should be easily correctable through adjustment of T_{60} and IACC in the model outputs. Results further suggest that spatial rendering using individualized HRTFs offers little benefit over nonindividualized HRTF rendering for room simulation applications where source direction is fixed.
 [DOI: 10.1121/1.3167842]

PACS number(s): 43.66.Lj, 43.55.Hy [RLF] Pages: 776–791

I. INTRODUCTION

Binaural technology has enabled realistic virtual listening simulation of a variety of room environments from anechoic rooms to concert halls. These "auralization" techniques not only allow users the unique opportunity to listen and evaluate the acoustics of different environments without being physically present in the environments but they also afford architectural acousticians, sound engineers, and scientists (among others) levels of control of the acoustic stimulus reaching the listeners' ears that would be impractical or perhaps even impossible in real acoustic listening spaces. Although many aspects of the complex methods underlying particularly model-based auralization techniques (Kleiner *et al.*, 1993; Vorländer, 2008) continue to be improved and refined (see Rindel, 2000), perceptual evaluation of the end results has received relatively little attention. Arguably the most severe form of perceptual testing for evaluating auralization methods would be to determine whether listeners could discriminate sound signals in a real room from virtual simulations designed to emulate the same source signal in the same room. For anechoic listening rooms, such tests have been conducted and under the best conditions of the simulation, real and virtual are indistinguishable (Zahorik *et al.*, 1995; Hartmann and Wittenberg, 1996; Kulkarni and Colburn, 1998; Langendijk and Bronkhorst, 2000). Analogous testing in more complicated reverberant room environments has yet to be conducted and for good reason. Best evidence suggests that even the most sophisticated model-based auralization techniques cannot reproduce the acoustic stimulus measured in a real room to less than the just-noticeable limits for human listeners on a host of room-acoustic parameters when estimated in isolation (Vorländer, 1995; Bork, 2000). Although this implies that even the best virtual room simulations would be discriminably different from real-room listening, the simulations are perhaps no less "natural" or "room-like" or different on perceptual properties that might depend on multiple naturally covarying physical parameters. Alternative methods are therefore needed to more fully assess the perceptual similarity of real and virtual room simulations. This article describes and implements one type of alternative method in which similarity ratings between both acoustically measured and modeled rooms are evaluated using multidimensional scaling (MDS) techniques. A principal advantage of this method over simple discrimination testing is that it allows the potentially multiple perceptual aspects or dimensions by which listeners rate the similarities of different measured or modeled room-acoustic simulations to be explicitly determined.

MDS techniques have been applied to a variety of problems in the hearing sciences, including the perception of vowels (Kewley-Port and Atal, 1989), consonant confusions (Bilger and Wang, 1976; Soli and Arabie, 1979), vocal qualities (Kempster *et al.*, 1991), timbre (Grey, 1977), and the perceptual properties of concert hall acoustics (Yamaguchi, 1972). In general, MDS techniques seek to determine a configuration of the experimental stimuli in a hypothetical Euclidean space that optimally describes or represents participants' judgments of similarity (or disimilarity) between all possible pairs of stimuli in the experiment. Stimulus pairs that are judged to be similar will lie close together in this

a)Electronic mail: pavel.zahorik@louisville.edu

"perceptual" space, and stimuli that are judged to be very different will lie far apart in the perceptual space. Different (independent) dimensions in the derived perceptual space can then be interpreted as the different perceptual attributes or quantities by which participants base their judgments. Further interpretation of the perceptual space dimensions is often accomplished by noting relationships in these dimensions to physical aspects of the stimuli. For example, work by Yamaguchi (Yamaguchi, 1972) concluded that listeners' judgments of similarity between various seating positions within two concert halls were based on three perceptual parameters, since the MDS solution for the listeners' similarity judgments was found to be three-dimensional. The first two dimensions of this solution were highly correlated with the physical parameters of sound pressure level and reverberation time, and therefore likely represent perceptual correlates of these parameters. The third dimension in the scaling solution was not easily interpretable in relation to any physical stimulus quantities. Although the precise relationship between physical aspects of concert hall acoustics and relevant perceptual aspects is an area of active study and debate, the pioneering results of Yamaguchi using MDS procedures are in many ways similar to other results using both related (Schroeder *et al.*, 1974) and relatively unrelated means for assessing perceived similarity or preference (Barron, 1988; Beranek, 2004).

In the work described here, MDS was used to assess the perceptual similarity between auralizations using both measurements from real rooms and simple room-acoustic models. The perceptual accuracy of the models is then reflected in their proximity in the MDS solution to stimuli based on measurements from real rooms. Given the results of past room modeling evaluations (Vorländer, 1995; Bork, 2000), some perceptual differences between measured and modeled stimuli are expected. Interpretation of the perceptual dimensions resulting from the scaling solutions will allow for more detailed assessment of the particular perceptual aspects in which the models depart from real rooms.

One obvious issue related to the MDS methods as described here is that substantial variability in similarity ratings from participant to participant might naturally be expected. Although classical MDS procedures offer no way of accounting for this variability, more recent weighted MDS procedures (e.g. INDSCAL; Carroll and Chang, 1970) allow the extent to which individual participants' responses are based on a given dimension in the stimulus-space to be determined. Each individual participant can then be characterized by the weight they place on each stimulus dimension. In this way, individual differences can be effectively analyzed. Although previous studies of perceptual similarity in room acoustics have not implemented methods to scale individual differences (Yamaguchi, 1972; Schroeder *et al.*, 1974), it seems clear that the potential for considerable individual differences exists in this application. As such, INDSCAL methods are implemented in the current study.

A variety of techniques for producing auralizations based on acoustic models of a room environment have been proposed (see Kleiner *et al.*, 1993; Rindel, 2000 for review) and implemented in commercially available software (e.g. ODEON and CATT-ACOUSTIC packages). Most techniques rely on assumptions of geometrical acoustics (Kuttruff, 2000), and many use separate methods for simulating early reflections and late reverberation. Because early reflections are typically more distinct both temporally and spatially than the late reverberant energy, which is more diffuse and homogeneous in time, they are modeled with more precise, and therefore more computationally demanding techniques. The modeling techniques implemented here adopt this same strategy, based loosely on methods described by Heinz (1993). An image-model (Allen and Berkley, 1979) is used to model early reflections and a statistical model is used for late reverberant energy modeling. Both individualized and nonindividualized head-related transfer functions (HRTFs) are used for spatially rendering the direct-path and early reflections. The end result of the model is an estimated binaural room impulse response (BRIR). Modeled BRIRs can then be compared to measured BRIRs, which are complete descriptions of the transfer characteristics of the various acoustical components of a given real listening situation including characteristics of the source, the room, and the listener's head and external ears. Overall, the model implemented here is rudimentary at best, and results in a variety of compromises related to the simulation of the acoustics of analogous real rooms. Areas of known compromise include appropriate simulation of source directionality, appropriate simulation of non-specular aspects of early reflections (i.e. scattering and diffraction), and appropriate simulation of late reverberant energy (correct effect of diffusion, etc.). No claims are therefore made as to the superiority of this modeling technique over other techniques. Nevertheless, the model is believed to maintain many of the essential perceptual aspects of small-room acoustics, while still allowing the experimenter complete control of all modeling methods and procedures—something that is often compromised in commercially available auralization packages.

The primary goal of this study is to evaluate the perceptual similarity of rooms simulated using these highly simplified modeling techniques to simulations based on measurements from a real room. As part of the evaluation, judgments of similarity will also be solicited for other rooms, with differing size, reverberant properties, and simulation fidelity. This will allow for determination of the acoustical factors that are most relevant for judgments of perceptual room similarity in this, and perhaps other similar sets of rooms, while at the same time providing a means for perceptual validation of the proposed room modeling techniques.

A secondary goal of this study is to determine the necessity of individualized HRTFs for realistic spatial rendering in the room simulation, which is one potentially important aspect of room simulation fidelity. Although individualized HRTFs are known to result in superior spatial rendering of sound source direction (Wenzel *et al.*, 1993), the process of measuring HRTFs for each potential user of an auralization system is a significant logistical difficulty. It is therefore important to carefully quantify the potential benefits of individualized HRTFs for room auralization applications.

The majority of past work both related to room modeling techniques and to the perceptual aspects of room acoustics has focused on concert hall environments, many of which have interior volumes of 20 000 m^3 or more. The current study, however, is concerned primarily with smaller room listening environments (between approximately 14 and 7800 m^3), which are more representative of everyday listening environments in which the vast majority of our auditory functioning takes place. Given that the acoustic contributions of a room are known to affect many critical auditory abilities, such as speech intelligibility (Peutz, 1971; Nabelek and Robinson, 1982) and sound localization in both direction (Hartmann, 1983) and distance (Zahorik, 2002), further understanding of the perceptual attributes of the acoustic properties themselves from everyday room environments is essential. In addition to addressing an important and generally understudied area, focus on small-room acoustics, particularly with simple rectangular shapes, has the methodological benefit of lessening the computational complexity in theory required for effective room modeling. It is important to note that the room modeling techniques implemented here become inappropriate in cases where wave behavior of sound in an enclosed space can no longer be ignored, such as when room dimensions and source/receiver distances become small relative to sound wavelength (Lam, 2005). Such behavior should be largely irrelevant for the situations examined in this study.

Related work by Berkley and Allen (1993) used classical MDS techniques to determine the perceptual similarity between 5 small rooms (constant interior volume of 75.5 m^3 with variable surface absorptions and source distances in each room), all simulated using an image model (Allen and Berkley, 1979) and presented monaurally without HRTF spatial rendering. Results from this important and highly relevant work suggest that listeners base their judgments of room similarity on two perceptual dimensions: one related to reverberation time and one related to variation in the sound spectrum. Although a number of methodological differences exist between Berkley and Allen's study (1993) and the work described here, their results will nevertheless serve as a basis for comparison of the results reported here in which model-based room simulations are compared both physically and perceptually to simulations based on measurements from a real room.

II. METHODS

A. Room-acoustic measurements and modeling

1. Participants

Nine listeners (6 female) ages 18–31 years participated in the acoustical measurement phase of this study.

2. BRIR measurements

BRIRs were measured for each participant in a single rectangular room using methods fundamentally identical to those described in Zahorik 2002. The room was large rectangular office room with dimensions of $5.7 \times 4.3 \times 2.6$ m^3 ($L \times W \times H$). Walls were painted drywall material. The floor was carpeted (short, dense weave), and the ceiling was a suspended type, constructed of acoustical tile materials. The participant was seated (1.3 m from floor to ear level) in the approximate center of the room: 3.8 m from the front wall, and 2 m from the left-hand side wall. All measurements were made using binaural microphones (Sennheiser KE4-211-2) placed at the entrance of the acoustically sealed ear canals (i.e. blocked-meatus configuration). Previous research has shown that this measurement configuration when paired with appropriate headphones and compensation will produce results vary similar to those obtained using probe-microphones placed near the tympanic membrane (Hammershøi and Møller, 1996). An additional benefit of this microphone configuration is that is allows larger microphones to be used, with frequency response and noise characteristics that are generally superior to probe-microphones. The sound source was a small full-range loudspeaker (Cambridge SoundWorks Center/Surround IV) with high-quality amplification (D-75, Crown, Inc.) positioned at ear level directly in front of the participant at a distance of 1.4 m Standard system identification techniques using a maximum-length sequence (MLS) signal (Rife and Vanderkooy, 1989) were used to measure BRIRs for each participant. The responses to a 16th order MLS signal (65535-sample) presented periodically were averaged coherently (ten averages) in order to improve signal-to-noise ratio, which was at least 55 dB (broadband) in all cases after averaging. Impulse responses were derived from the averaged responses via circular cross-correlation (Rife and Vanderkooy, 1989). All signal generation and data acquisition was performed using MATLAB software (Mathworks, Inc.), and high-quality D/A and A/D hardware (DD1, Tucker-Davis Technologies, Inc.) using 16-bit quantization and a 48 kHz sampling frequency. No compensation for the response characteristics of the loudspeaker was applied to the measurements. The loudspeaker was relatively omni-direction up to approximately 1 kHz, as is evident in its directional response data (Fig. 1) measured using procedures detailed in ISO-3382 (ISO-3382, 1997). Additional details of the measurement room are shown in Table I, along with parameters for subsequent physical and psychophysical testing including whether the BRIR measurements originated from the listener's own ears (ID 1) or from another participant's ears (IDs 2–3).

3. BRIR models

Simple models of BRIRs were constructed using a three-dimensional image-model (Allen and Berkley, 1979) to simulate early specular reflections within a hypothetical rectangular room and a statistical model of the late diffuse reverberant energy. This approach, while relying on a variety of assumptions and simplifications, is fundamentally similar to that described by Heinz (1993), and is shown schematically in Fig. 2. Twelve different listening situations ("rooms") were modeled, all with an omni-directional sound source directly in front of the listener's location in the approximate center of each room, at a distance of 1.4 m, and 1.3 m above the floor surface (approximately ear level when seated). The rooms differed in various physical parameters such as size and the amount of surface absorption, as well as other details of the simulation methods, which are described

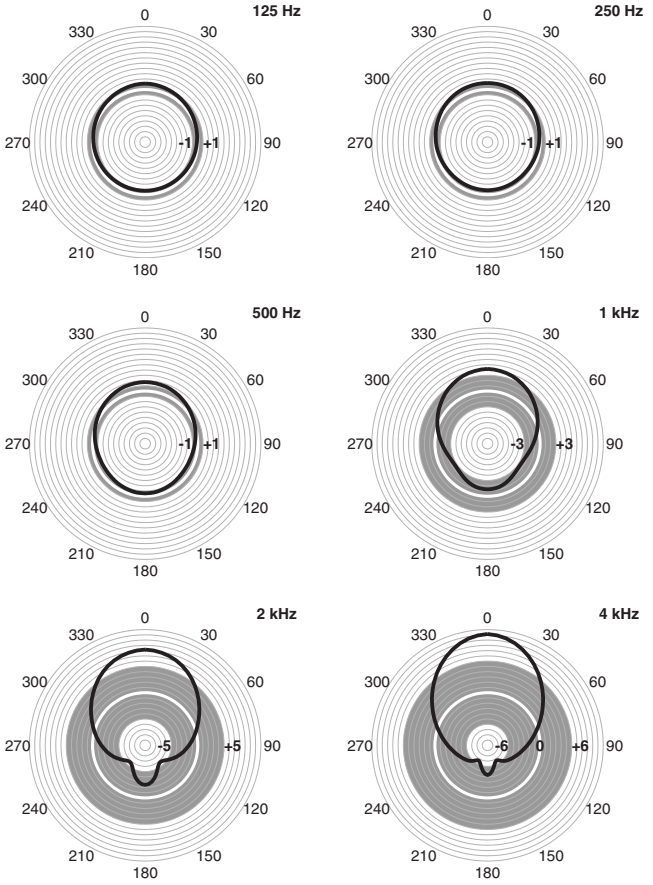


FIG. 1. Directional responses of the measurement loudspeaker (Cambridge SoundWorks Center/Surround IV) in octave bands (125–4000 Hz) expressed in decibels (1 dB/division) relative to the 360° (energetic) average response (0 dB, white curve). The 0° orientation indicates that the loudspeaker was oriented directly facing the (omni-directional) measurement microphone. Shaded tolerance regions for source omni-directionality from ISO-3382 (1997) are shown for each octave band. Below 2 kHz, the measurement loudspeaker deviates from omni-directional specifications by at most 1.2 dB. Above 2 kHz, the directional characteristic of this loudspeaker becomes more pronounced, but still only deviates from omni-directionality by at most 5.2 dB at 4 kHz.

below and in Table I. Certain room simulations were designed to closely approximate the measurement room environment in which the BRIR measurements described in Sec. II A 2 were conducted (e.g., IDs 4–6). Other room simulations were designed for comparative use in subsequent psychophysical scaling experiments. For example: room 15 was anechoic, and room 14 had only early reflections and no late reverberant energy, rooms 7–10 differed in size, and rooms 11–13 differed in surface absorption.

a. Early response modeling. The direct-path and 500 early reflections were all spatially rendered with head-related transfer functions (HRTFs) measured using techniques similar to those described by Wightman and Kistler (1989). HRTFs were measured for each participant from a spherical grid of 541 spatial locations surrounding the listener (10° spacing, full 360° in horizontal angle; vertical angles from −60°. below to 90° above ear level) in an anechoic chamber using miniature electret microphones (Sennheiser KE4-211-2) in a blocked-meatus configuration. The spherical grid of measurements was conducted using vertically-oriented semi-circular (1.4 m radius) array of 16 loudspeakers that could be rotated horizontally around the participant's head location at the center of the grid. A given spatial location in the grid was selected by rotating the arc to the appropriate horizontal (azimuth) angle, and then energizing the loudspeaker on the arc corresponding to the appropriate vertical (elevation) angle. Arc rotation was accomplished using a high-torque computer-controlled motor (model HA5C, Haas, Inc.) with 0.01° rotational precision. Loudspeaker switching was performed prior to audio signal amplification (D-75, Crown, Inc.) using a computer-controlled switching device (AM-16/B, 360 Systems, Inc.). The measurement signal was a 20.48 ms broad band (0.2–25 kHz) noise constructed with a phase-spectrum that minimized the peak-factor of the signal (Schroeder, 1970). For each measurement location, 100 repetitions of this signal were presented periodically at a level producing 70 dBA at the recording location (center of the participant's head). The signals were presented and the responses from the binaural microphones were recorded using the same high-quality D/A-A/D hardware (DD1, Tucker-Davis Technologies, Inc.) at a sampling frequency of 100 kHz, with 16-bit precision. Responses were averaged coherently (100 averages) in order to improve signal-to-noise ratio. Transfer-functions were derived for each measured response via frequency-domain division by the measurement signal. Results for each measurement location were then down-sampled to 50 kHz and windowed to 1024-points in order to facilitate efficient storage and later convolution operations. Using these techniques, a set of 541 measurements could be completed in approximately 30 min.

These "HRTF" measurements differed from most standard types of HRTF measurements in one important respect: They were not referenced relative to the measured response from a reference microphone in the absence of the head. This was done in order to preserve the response of the measurement loudspeakers in each of the transfer function measurements, for later comparison with the room BRIR measurements which also contained the response of the loudspeaker (same make and model).

Reflection locations determined from the image-model were rendered to the nearest HRTF measurement angle. No interpolation was implemented. Individualized HRTFs were used for spatial rendering in certain simulation conditions (e.g. simulation IDs 4 and 7–15). Nonindividualized HRTFs from different human participants were used in other conditions (e.g., rooms 5 and 6). All reflections were modeled as ideal specular reflections resulting from a point-source (omni-directional), with no frequency-dependent absorption characteristics and no dependencies on angle of incidence. Broadband levels of each reflection were determined based on path-length (r), an average broadband energy absorption coefficient (α) for all surfaces in the room, the reflection order (n), and a constant loss factor (f) designed to help offset level discrepancies due to scattering and/or other factors. The gain for the jth reflection relative to the direct-path (with pathlength r_0) was

TABLE I. Description of measured (IDs 1–3) and modeled (IDs 4–15) BRIRs evaluated in this study. For each BRIR, dimensions (length × width × height), volume (V), surface area (SA), and source distance (d) are listed. Individualized and nonindividualized HRTF implementations are also indicated (subject ID code listed for nonindividualized sets), along with the number of HRTF sets (n) used for generating the full stimulus set in each room. For the modeled BRIRs, early and late energy absorption coefficients (α) entered into the model are also indicated.

											Late α					
ID	HRTF set	n	L (m)	W (m)	H (m)	V (m^3)	SA (m^2)	d (m)	Early α	T_e (ms)	125	250	500	1000	2000	4000
Meas.																
1	Indiv.	7	5.7	4.3	2.6	62.4	99.6	1.4	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
2	SXB	1	5.7	4.3	2.6	62.4	99.6	1.4	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
3	SZM	1	5.7	4.3	2.6	62.4	99.6	1.4	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
Modeled																
4	Indiv.	7	5.7	4.3	2.6	62.4	99.6	1.4	0.29	55	0.40	0.30	0.30	0.30	0.22	0.20
5	SXB	1	5.7	4.3	2.6	62.4	99.6	1.4	0.29	55	0.40	0.30	0.30	0.30	0.22	0.20
6	SZM	1	5.7	4.3	2.6	62.4	99.6	1.4	0.29	55	0.40	0.30	0.30	0.30	0.22	0.20
7	Indiv.	7	3.4	2.6	1.5	13.5	35.9	1.4	0.29	34	0.40	0.30	0.30	0.30	0.22	0.20
8	Indiv.	7	8.5	6.4	3.9	210.5	224.1	1.4	0.29	83	0.40	0.30	0.30	0.30	0.22	0.20
9	Indiv.	7	11.3	8.5	5.2	499.0	398.4	1.4	0.29	112	0.40	0.30	0.30	0.30	0.22	0.20
10	Indiv.	7	28.4	21.3	12.9	7797.4	2490.0	1.4	0.29	283	0.40	0.30	0.30	0.30	0.22	0.20
11	Indiv.	7	5.7	4.3	2.6	62.4	99.6	1.4	0.05	55	0.05	0.05	0.05	0.05	0.05	0.05
12	Indiv.	7	5.7	4.3	2.6	62.4	99.6	1.4	0.10	55	0.10	0.10	0.10	0.10	0.10	0.10
13	Indiv.	7	5.7	4.3	2.6	62.4	99.6	1.4	0.30	55	0.30	0.30	0.30	0.30	0.30	0.30
14	Indiv.	7	5.7	4.3	2.6	62.4	99.6	1.4	0.29	55	1.00	1.00	1.00	1.00	1.00	1.00
15	Indiv.	7	5.7	4.3	2.6	62.4	99.6	1.4	1.00	55	1.00	1.00	1.00	1.00	1.00	1.00

$$20\log_{10}\left[\frac{r_0\sqrt{1-f\alpha}}{r_j n_j}\right]\ \text{dB}. \tag{1}$$

For certain models, average α was estimated based on published α values for common building materials (Moulder, 1991) averaged across frequency octave-band frequencies from 125 to 4000 Hz and weighted by the relative surface area of each material in the modeled room. Other models used experimentally altered values for average α. Table I displays the specific choices for HRTF-individualization and (early response) α values for the simulated rooms evaluated in this study. Also displayed in Table I is the delay relative to the direct-path of the last image (500th) in the estimates of the early response, T_e. The constant loss factor, f, was determined via pilot testing and set to a value of 3 for all models. A schematic early response is shown in Fig. 2(a).

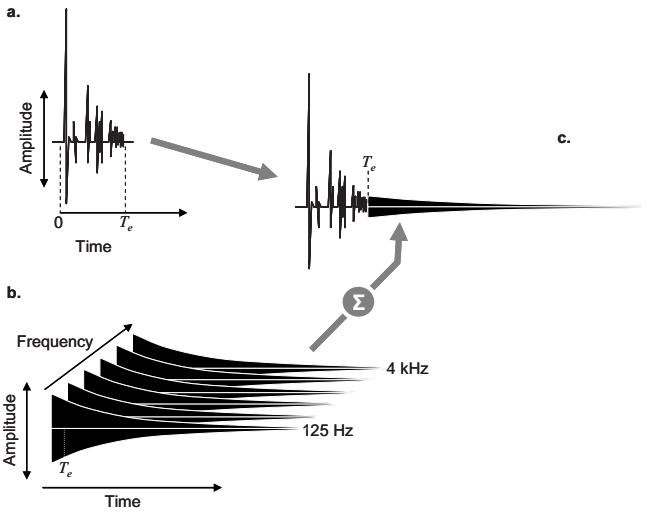


FIG. 2. Schematic representation of the BRIR modeling procedure. (a) Stylized representation of the early portions of the BRIR (one ear only) estimated using an image model (Allen and Berkley, 1979) and then spatially rendered using HRTFs. The time axis represents the duration of the early response, T_e, relative to the start of the direct-path response (0). (b) Stylized representation of the late response modeling procedure in which noise samples shaped by exponentially decaying envelopes are generated in each of six octave bands (125 Hz–4 kHz) and then summed and scaled (see text for details). (c) Representation of the combined early and late responses for one ear. Identical procedures were implemented to estimate the response of the other ear.

b. Late response modeling. Diffuse late reverberation was simulated using independent Gaussian noise samples for each ear shaped by separate decay functions applied to each of six octave-bands ranging from 125–4000 Hz [see Fig. 2(b)]. The decay functions were derived from the Sabine equation (Sabine, 1922):

$$T_{60} = 0.163\frac{V}{S\bar{\alpha}_i}, \tag{2}$$

which estimates the amount of time (s) required for sound level to decay by 60 dB (T_{60}) following the offset of a source signal, from the parameters of room volume (V), total surface area of the reflecting surfaces (S), and the average absorption coefficient for all surfaces within the ith octave band ($\bar{\alpha}_i$). Average absorption coefficients in each band were again based on published α values for common building materials (Moulder, 1991) weighted by the relative surface area of each material. These estimated T_{60} values were used to define decay functions for each octave band of the following form:

$$d(t) = 10^{-3t/T_{60}}, \quad (3)$$

where t is measured in seconds. The parameters used to estimate T_{60} and compute all decay functions for modeling late reverberant responses are shown in Table I. Broadband late responses were created by summing the decay-shaped noise samples across octave bands.

c. Combining early and late responses. The levels of early and late responses were first matched by noting the rms level (broadband) in the last 10 ms of the early response (i.e. $T_e - 10$ ms to T_e) and then scaling the late response such that its rms level (broadband) over the same period ($T_e - 10$ ms to T_e) was identical to the early response. All energy in the late responses between 0 and T_e was then removed, and the resulting late response for each ear was summed with the early response for each ear to create an estimated BRIR, as shown schematically in Fig. 2(c). The resulting BRIR was then down-sampled to 48 kHz and stored for subsequent analysis and psychophysical testing. All room model and signal processing was implemented using MATLAB® (Mathworks, Inc.) software.

4. Equalization for headphone presentation

In order to facilitate accurate reproduction of the appropriate pressure waveform at both eardrums using headphone-based virtual auditory space techniques, the transfer characteristics of an acoustically open headphone (Beyerdynamic DT 990 Pro) when coupled to the head were measured for each participant. These measurements were obtained during the same measurement sessions as the BRIR and anechoic HRTF measurements using the same binaural microphones (Sennheiser KE4-211-2) in a blocked-meatus configuration and similar techniques for system identification. Results of these measurements were used to construct headphone equalization filters to correct for the response of the headphone when coupled to the head of each participant, following logic described and evaluated by Møller, Hammershøi, and colleagues (Møller, 1992; Møller *et al.*, 1995). Because the equalization quality can depend on the degree to which the microphone position in the ear canal was similar for both headphone and HRTF or BRIR measurements, two sets of equalization filters were made for each participant: one based on headphone measurements during the BRIR measurement session and the second based on measurements during the HRTF session. The former equalization filters were then used for virtual room simulation using the measured BRIRs, and the later equalization filters were used for model-based room simulation. Methods to construct the equalization filters were similar to those described in previous work (Zahorik, 2002). The magnitude spectrum from each measurement was inverted, smoothed (20% of a critical bandwidth), and low-pass filtered at 20 kHz. The results were then defined to have linear phase, and used to implement a 256-coefficient finite impulse response filter for headphone equalization (48 kHz sampling rate).

B. Physical testing

Measured and modeled BRIRs were evaluated physically using two general methods: one in which the BRIRs were directly compared between the measured room and the best-case model, and one in which various common room acoustical parameters were computed from the BRIRs for each room. Parameter values were then compared and used to provide a basis for interpretation of subsequent psychophysical testing.

1. BRIR comparisons

The direct analysis of BRIR similarity compared BRIRs from the measured room (ID 1) to those from the best-case modeled room (ID 4) for each of nine participants. Comparisons were made by first bandpass filtering (third-order Butterworth as specified by ANSI-S1.11, 2004) the BRIRs (left and right ears separately, un-equalized for headphone reproduction) into 1/3 octave bands, with center frequencies ranging from 125 to 8000 Hz. In each band, the normalized cross-correlation function, CF, was computed between measured and modeled BRIRs:

$$\mathrm{CF}_{L \text{ or } R}(\tau) = \frac{\int_0^{t_{\max}} p_1(t) p_4(t+\tau) dt}{\sqrt{\int_0^{t_{\max}} p_1^2(t) dt \int_0^{t_{\max}} p_4^2(t) dt}}. \quad (4)$$

CF was computed for each ear separately, where p_1 and p_4 are the FIRs from rooms 1 and 4, respectively, for a given ear (L or R). A variable maximum integration time, $t_{\max}$, which represents the maximum integration time applied to the impulse responses was also implemented. Here three different values of $t_{\max}$ were evaluated: 5 ms, 20 ms, and full impulse response, which will be denoted as $t_{\max}=\infty$. These different choices of $t_{\max}$ where chosen to determine how the degree of match may be influenced by the direct-path alone ($t_{\max}=5$ ms), the inclusion of early reflections ($t_{\max}=20$ ms), and the inclusion of early reflections as well as late reverberant energy ($t_{\max}=\infty$). Each 1/3-octave-band CF was then summarized by computing the cross-correlation coefficient, CC, defined as

$$\mathrm{CC} = \max|\mathrm{CF}(\tau)| \quad \text{for } -1 \text{ ms} < \tau < 1 \text{ ms}, \quad (5)$$

which is simply the maximum magnitude of CF CC may be interpreted as the degree of linear association between the two impulse responses that is independent of delay or polarity. High similarity between measured and modeled BRIRs will yield CC values near 1.

2. Acoustical parameter comparisons

Five different room-acoustic parameters were estimated from the measured and modeled BRIRs in this study. The majority of the estimated parameters were commonly used room-acoustic parameters (ISO-3382, 1997), including reverberation time (T_{60}), clarity index (C_{50}), center time (T_c), and the interaural cross-correlation coefficient (IACC) based on full-duration BRIRs. An additional spectral centroid parameter, f_c, not described in ISO-3382 (1997) was also estimated in an attempt to characterize any potential timbral differences between the BRIRs. All parameters were estimated from the BRIRs for each measured/modeled room (un-equalized for headphone reproduction) for each participant.

Estimation procedures for T_{60}, C_{50}, T_c, and IACC were based on those described in ISO-3382 (1997) with the following important differences: (1) All BRIRs in this study resulted from measurements made with directional microphones placed in the ears of individual listeners. Although this technique is valid for estimation of IACC, ISO-3382 (1997) recommends that estimates of T_{60}, C_{50}, T_c be made from measurements with an omni-directional microphone in the absence of the head. Here, these parameters were instead estimated from the left ear portion of the BRIRs. (2) ISO-3382 (1997) also requires that an omni-directional source be used for all parameter estimation. The source used for all measured BRIRs in this study had directional response properties that deviated from true omni-directionality. Although both of these departures from ISO-3382 (1997) recommendations may have biased the parameter estimates reported in this study, the bias due to directional microphones used in the measurements should be relatively constant across all measurements, and the bias due to source directionality in the measure BRIRs is believed to be relatively low below 2 kHz (see Fig. 1).

Estimation of the spectral centroid parameter, f_c, for each BRIR was accomplished as follows. Each BRIR was first passed through a bank of 1/6th-octave rectangular bandpass filters, with center frequencies, cf_i, ranging from 125 to 16 000 Hz. Let E_i be the resulting energy in the ith 1/6th octave band specified in decibels. The spectral centroid, f_c, in hertz is therefore defined as

$$f_c = 10^{(\Sigma_{i=1}^{n}(\log 10(\mathrm{cf}_i)E_i)/\Sigma_{i=1}^{n}(E_i))}, \quad (6)$$

where $n=43$ in this case, corresponding to the number of bandpass filters used in the analysis. Conceptually, the spectral centroid is the center of mass of a signal's magnitude spectrum, and has been shown to be related to the perceptual quality of timbre (Grey and Gordon, 1978). Since previous work has identified the importance of spectral/timbral aspects in small-room acoustics (Bech, 1995, 1996), and informal observation suggests that mismatches in HRTF processing can cause changes to the timbre of reproduced sound, this spectral centroid parameter may be particularly relevant for the listening situations examined in this study.

Due to specific experimental and measurement procedures of this study, three additional and common room-acoustic parameters (all described in ISO-3382, 1997) were not estimated in this study: sound strength, lateral energy fraction, and early decay time. Sound strength, which is a measure of sound energy in a given room relative to energy at a fixed source-receiver distance (typically 10 m) in the acoustic free-field, was not estimated here because overall sound presentation level was equalized across all sound stimuli in this study. This caused sound strength to be essentially fixed, and therefore not a useful acoustic parameter for describing acoustical differences across the measured and simulated rooms in this setting. Lateral energy fraction, which is the proportion of laterally arriving sound energy relative to omni-directional energy, was also not estimated, since this parameter requires measurements from a figure-eight microphone which was not available for this study. Finally, early decay time was not separately estimated because preliminary testing revealed that it was almost perfectly correlated with measures of T_{60} for the rooms examined in this study.

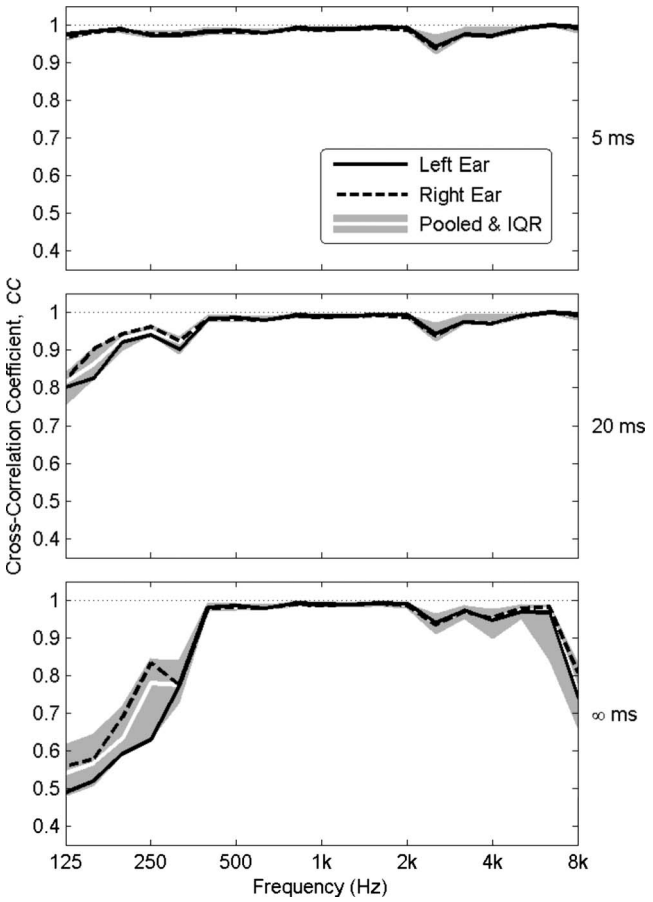


FIG. 3. Results of a correlation analysis between measured (ID 1) and best-case modeled (ID 4) BRIRs for all participants ($n=9$). The dependent measure is a cross-correlation coefficient, CC, which is the maximum magnitude of the cross-correlation function, $\mathrm{CF}(\tau)$ between measured and modeled impulse responses computed separately in each 1/3-octave band ranging from 125 to 8 kHz (see text for details). Three different maximum integration times for the cross-correlation function are shown: 5, 20, ∞ ms. In each plot panel, the median CC across all participants is displayed for left ear responses only (solid curve), right ear responses only (dashed curve), and for responses pooled across both ears (white curve). Shaded regions indicate the IRQ for all CC values.

C. Psychophysical testing

1. Participants

Seven listeners (six female) ages 18–31 years participated in the experiment. All had normal hearing, as verified by standard (ANSI-S3.9, 1989) audiometric screening at 15 dB HL from 125 to 8000 Hz, and were experienced in sound localization tasks. All listeners participated in the previous physical measurement phase of this study. Listeners SXB and SZM (see Table I) did not participate in this phase of the study.

2. Stimuli and presentation apparatus

15 different stimuli were constructed based on the 15 different measured or modeled room simulations detailed in Table I. The source signal for all stimuli was a high-quality

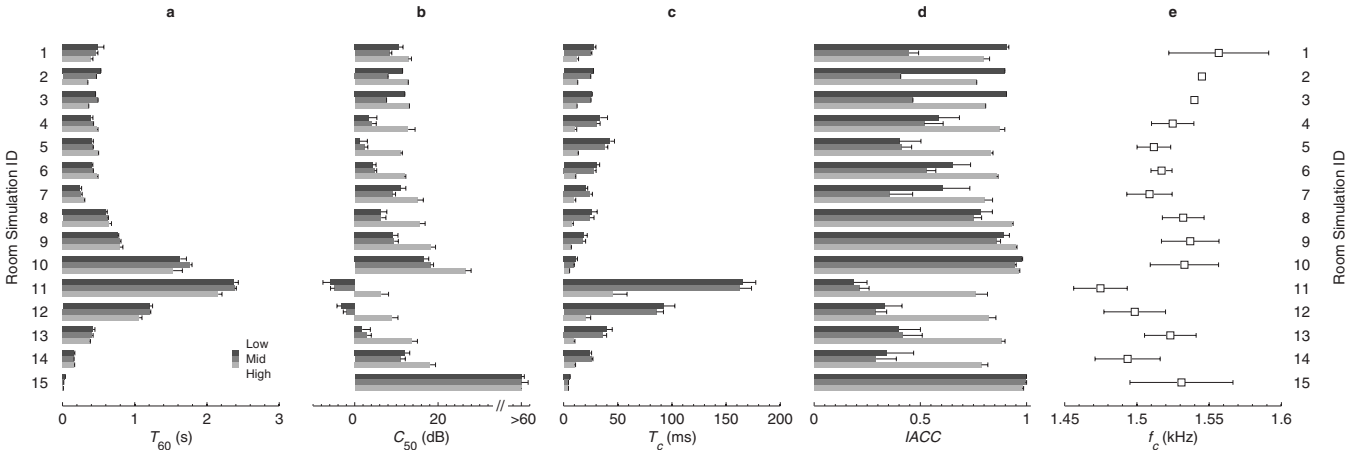


FIG. 4. Estimated room-acoustic parameters for 15 BRIR types (see Table I for ID coding and descriptions): reverberation time T_{60} (a), clarity index C_{50} (b), center time T_c (c), interaural cross-correlation IACC (d), and spectral centroid f_c (e). Separate estimates are shown for low (125 and 250 Hz bands), medium (500 and 1000 Hz bands), and high (2 and 4 kHz bands) frequency ranges in (a)–(d). In all cases, the displayed estimates are means across participants. Error bars represent one standard deviation.

speech sample (3.4 s duration) from a male talker recorded in anechoic space. This signal was convolved with BRIRs from each room. Overall level of the convolved stimulus was then equalized across all stimuli by matching the rms amplitudes. This was done in an attempt to remove overall level as a potential means for perceptually classifying the stimuli. Stimuli were presented in a double-walled sound booth over equalized headphones (Beyerdynamic DT-990-Pro) using Tucker-Davis Technologies equipment for D/A conversion (16-bit, 48 kHz) and headphone amplification (fixed gain). All signal processing was implemented using MATLAB® (Mathworks, Inc.) software.

3. Design and procedure

Participants listened to all possible pairs of different stimuli (210 total), presented with an inter-stimulus interval of 1 s. Participants were told to rate the perceived disimilarity between each stimulus in the pair using a 100-point rating scale, ranging from 0="exact same" to 99 ="completely different." Participants were allowed to listen to the stimulus-pair as many times as they wished prior to making their rating response, which the participant entered numerically on a computer keypad. No feedback was given to participants as to the type of trial or the nature of their responses. The experiment was run in blocks of 210 trials consisting of one set of all possible pairs of different stimuli, presented in random order. Participants required approximately 45 min to complete one trial block. Each listener completed nine blocks of trials, resulting in a total of 1890 trials, or nine similarity ratings for each stimulus pair. Listeners were explicitly instructed to note any stimuli in which the sound source location was not perceived external to the head.

III. RESULTS

A. Physical testing

Overall, the simple virtual room modeling techniques described in this study produced a reasonably good physical match to the measured room (a large office space). Quantitative assessment of the degree of physical matching was conducted both via direct analysis of BRIR similarity and via an indirect analysis of various room-acoustic parameters derived from the BRIRs. Results from these analyses are reported in Secs. III A 1 and III A 2.

1. BRIR comparisons

Figure 3 displays results from the BRIR correlation analysis comparing a measured BRIR to the best-case modeled BRIR. When t_{max}=5 ms, the cross-correlation coefficient was greater than 0.93 at all frequencies. This high degree of association was not particularly surprising, given that this time range was dominated by the direct-path response, which should have been very similar in both cases (e.g. same participant, same source, and direction). When t_{max}=20 ms, some decrease in correlation (mismatch) may be observed at frequencies below 400 Hz. This effect becomes more pronounced at $t_{max}=\infty$, and additional decreases in correlation may be observed above 6 kHz. Overall, this analysis suggests that the modeled BRIRs are in good agreement with the measured BRIRs within the 400–6000 Hz bandwidth, but show increasing mismatch above and particularly below this frequency range when the full BRIRs are analyzed ($t_{max}=\infty$). Increased mismatch is also accompanied by increased variability across participants [i.e., greater interquartile range (IRQ)].

2. Acoustical parameter comparisons

Estimates of five acoustical parameters derived from the BRIRs from each room are displayed in Fig. 4. Results for the T_{60}, C_{50}, T_c, and IACC parameters [Figs. 4(a)–4(d)] were condensed into three two-octave frequency ranges: low (125 and 250 Hz bands), medium (500 and 1000 Hz bands), and high (2 and 4 kHz bands), following methods described in ISO-3382 (1997). This process was performed separately for each participant in a given measured or modeled room. Figure 4(e) displays the spectral centroid parameter for each measure/modeled room. All values displayed in Fig. 4 repre-

TABLE II. Broadband acoustical parameters estimates (means across all frequency bands) for measured (ID 1) and best-case modeled (ID 4) BRIRs. Differences (and % changes) are also shown, as are estimates of the JND for each acoustical parameter determined from previous studies.

		Measured (ID 1)	Modeled (ID 4)	Difference		JND	
Broadband Parameter	T_{60} (ms)	445.8	432.3	13.5	(3%)	24 ms	(Seraphim, 1958)
	C_{50} (dB)	10.6	6.8	3.8	(36%)	1.1 dB	(Bradley *et al.*, 1999)
	T_c (ms)	21.8	24.9	−3.1	(−14%)	5.7–11.4 ms	(Cox *et al.*, 1993)
	IACC	0.718	0.659	0.1	(8%)	5%	(Okano, 2002)
	f_c (Hz)	1557	1525	32	(2%)	7%	(Emiroglu and Kollmeier, 2008)

sent mean estimates across participants. Standard deviation values across participants are also displayed (error bars).

Overall, considerable variation in the parameter estimates may be observed across the sample of rooms. This variation across rooms is generally much larger than the between-participant variation, with the exception of the f_c parameter where large individual variability was observed. This individual variability in f_c is likely due to differences in the HRTFs across participants, which is controlled through the use of individualized HRTFs in subsequent psychophysical testing. Two additional points regarding individual variability are noteworthy. First, there was zero individual variability between rooms 2 and 3, since each room was based on measured BRIRs from only a single participant. Second, although rooms 5 and 6 were also based on anechoic HRTF measurements from single participants, these rooms had nonzero parameter standard deviations. This is because the room models, which used independent samples of Gaussian noise for simulating the late reverberant energy, were re-computed for each participant in the study, even though the HRTF set was held constant. Hence the variability in these parameters was due solely to variability in late reverberant energy, not HRTF differences.

To further assess the adequacy of the room modeling procedures, parameter estimates between the measured room (ID 1) and the best-case modeled room (ID 4) were compared. Table II displays broadband parameter estimates for each room, and parameter estimate differences between rooms. Estimates of the just-noticeable difference for each parameter in isolation are also displayed based on results from previous studies that most closely approximated the listening conditions present in this study. Only the differences in C_{50} and IACC were greater than 1 just noticeable difference (JND), suggesting that rooms 1 and 4 would likely be discriminable based differences in either of these parameters alone.

An additional and important aspect of the room-acoustic parameter estimates for this set of rooms is that many parameters are highly correlated. These relationships are visible in Table III, which displays the Pearson correlation matrix between all pairs of parameters. Statistically significant correlation may be observed in many of the cells, particularly between the different frequency ranges for a given parameter. Significant correlations are also observed across parameters, particularly in relationship to the T_c parameter which appears to be highly correlated with many of the other parameters.

In an attempt to reduce the redundancy and dimensionality of this set of parameters, a principle-components analy-

TABLE III. Correlation matrix between physical room acoustic parameters for 15 BRIR types (see Table I for BRIR descriptions).

		T_{60}			C_{50}			T_c			IACC			
		Low	Mid	High	Low	Mid	High	Low	Mid	High	Low	Mid	High	f_c
T_{60}	Low	1.000												
	Mid	0.998[a]	1.000											
	High	0.990[a]	0.993[a]	1.000										
C_{50}	Low	−0.396	−0.392	−0.426	1.000									
	Mid	−0.331	−0.330	−0.355	0.980[a]	1.000								
	High	−0.296	−0.295	−0.317	0.968[a]	0.998[a]	1.000							
T_c	Low	0.757[a]	0.726[a]	0.738[a]	−0.498	−0.373	−0.341	1.000						
	Mid	0.753[a]	0.723[a]	0.735[a]	−0.493	−0.372	−0.341	0.998[a]	1.000					
	High	0.737[a]	0.704[a]	0.707[a]	−0.456	−0.359	−0.340	0.971[a]	0.977[a]	1.000				
IACC	Low	−0.287	−0.266	−0.309	0.582[b]	0.463	0.429	−0.704[a]	−0.708[a]	−0.618[b]	1.000			
	Mid	−0.116	−0.083	−0.081	0.644[a]	0.618[b]	0.627[b]	−0.582[b]	−0.592[b]	−0.610[b]	0.749[a]	1.000		
	High	−0.106	−0.071	−0.049	0.509	0.533[b]	0.562[b]	−0.483	−0.498	−0.589[b]	0.495	0.926[a]	1.000	
	f_c	−0.419	−0.407	−0.448	0.329	0.215	0.178	−0.703[a]	−0.722[a]	−0.647[a]	0.873[a]	0.537[b]	0.352	1.000

[a]Correlation is significant at the 0.01 level (two-tailed).
[b]Correlation is significant at the 0.05 level (two-tailed).

TABLE IV. Rotated principal component matrix for acoustic parameters estimated from 15 measured and modeled room simulations (see Table I for room descriptions). The strongest loadings for each component are indicated in bold.

		Component			
		1	2	3	4
T_{60}	Low	**0.975**	−0.159	−0.101	−0.009
	Mid	**0.972**	−0.169	−0.090	0.037
	High	**0.961**	−0.198	−0.148	0.078
C_{50}	Low	−0.243	**0.916**	0.232	0.185
	Mid	−0.173	**0.959**	0.099	0.198
	High	−0.141	**0.959**	0.057	0.234
T_c	Low	0.720	−0.118	−0.489	−0.437
	Mid	0.715	−0.113	−0.499	−0.447
	High	0.713	−0.087	−0.376	−0.566
IACC	Low	−0.147	0.312	0.886	0.275
	Mid	0.019	0.448	0.447	0.764
	High	−0.008	0.352	0.165	**0.899**
	f_c	−0.316	0.037	**0.901**	0.162

sis with varimax rotation was performed on the mean parameter estimates shown in Figure 4. From this analysis (implemented using SPSS® software), four principle components were found to account for 97.8% of the variance in the full set of parameter estimates (13 parameters for each of 15 room simulations). This suggests that the original 13-dimensional parameter space with a relatively high degree of redundancy can be effectively represented with only 4 dimensions. Components 1–4 accounted for 35.3%, 24.8%, 19.7%, and 18% of the variance, respectively, suggesting relatively equal contributions of components 1–4. The rotated component loadings resulting from this analysis are shown in Table IV, where each value may be interpreted as the correlation between a given component and room-acoustic parameter. From these results, it appears that component 1 is most strongly associated with T_{60} (relatively independent of frequency), component 2 is most strongly associated with C_{50} (also relateively independent of frequency), component 3 is most strongly associated with f_c, and component 4 most strongly associated with high-frequency IACC. This suggests that broadband T_{60}, broadband C_{50},f_c, and high-frequency IACC make up a set of four independent physical room-acoustic parameters for this set of room simulations.

Consistent with correlation results shown in Table III where T_c is shown to be highly correlated with many other room-acoustic parameters, T_c also does not appear to be strongly related to any principle component in this analysis. Instead, T_c is only moderately related to the principal component 1, which is the same component to which T_{60} is much more strongly related. This result, in conjunction with the correlation results (Table III), suggests that T_c is not independent of T_{60}, and that T_{60} is a better predictor of one source of independent variability in the acoustic parameters for the rooms in this study. For this reason the T_c parameters will be excluded from subsequent analyses.

B. Psychophysical testing

A MDS analysis was performed on the mean perceived similarity ratings (interval measurement scale assumed) from each listener between all possible pairs of measured/modeled room simulations listed in Table I. This analysis (INDSCAL, implemented using SPSS® software) allowed for a scale of perceived room similarity to be determined, as well as a characterization of each individual listener's use of the resulting scale. Here, the scale of perceived similarity in room acoustics was determined to be two-dimensional (R^2 =0.830), since solutions with higher dimensionality did not account for a substantially greater proportion of the total variance ($0.871 \leq R^2 \leq 0.889$, for solutions with three to six dimensions). The two-dimensional solution also resulted in high R^2 values associated with each individual listener's data, as shown in Fig. 5. There were no reports of non-

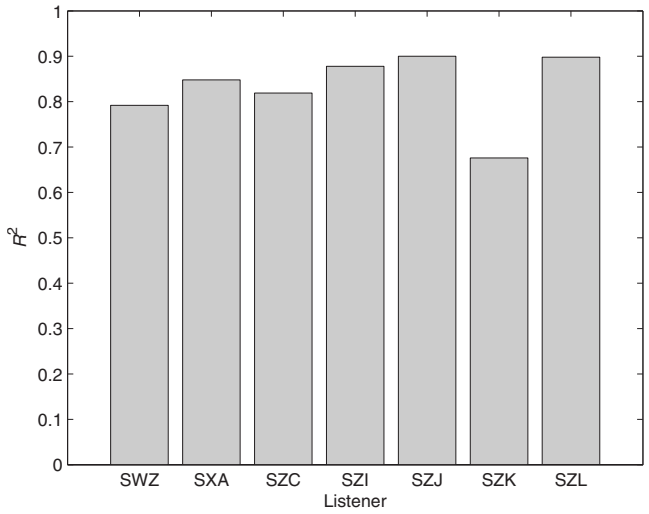


FIG. 5. Proportion of variance accounted for (R^2) in a two-dimensional scaling solution as a function of individual listener.

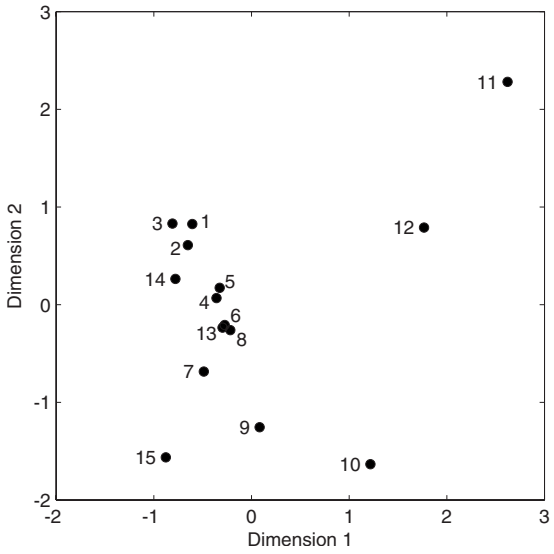


FIG. 6. Derived two-dimensional scale (R^2=0.830) of perceived room similarity (see Table I for room simulation ID coding and descriptions).

externalized sound sources from any of the listeners.

Figure 6 displays the resulting two-dimensional scale of perceived room acoustics. To further interpret the scale, each dimension in the scale was correlated with physical room-acoustic parameters that were found to be most representative of the acoustical variability between rooms in this study. Table V displays the correlation between each of the two dimensions in the psychophysical scale with four "most representative" physical parameters: those found to produce factor loadings with magnitude ≥0.9 in any of the four dimensions from the principal components analysis described in Sec. III A (see Table IV). Because Dimension 1 from the scaling solution was highly correlated ($r>0.94$) with T_{60}, at all frequencies, it likely represents a perceptual quantity related to reverberation time. The relationship is direct: Increases in Dimension 1 scale values correspond to increases in T_{60} parameter values. Dimension 1 also appears to be more moderately related (inversely) to the f_c parameter ($r=-0.58$), which may suggest that this perceptual dimension may also depend somewhat on sound timbre. Dimension 2 of the scaling solution appears to be most strongly related to the IACC parameters in the mid- and high frequencies ($r<-0.82$). As a result, it may represent a spatial aspect of the perceived sound, perhaps related to image size or diffuseness. The relationship is inverse, however: Increases in Dimension 2 scale values generally correspond to decreases in mid- and high-frequency IACC. More moderate (but still statistically significant) negative correlations between Dimension 2 and the C_{50} parameters as well as low-frequency IACC are present.

TABLE V. Correlation between physical stimulus parameters and values in the two-dimensional perceptual scaling solution. The T_c parameter was excluded from this comparison, because it was found to be highly correlated with T_{60} (see Table III).

		Dim. 1	Dim. 2
	Dim. 1	1.000	0.355
	Dim. 2	0.355	1.000
T_{60}	Low	0.946[a]	0.383
	Mid	0.943[a]	0.344
	High	0.948[a]	0.330
C_{50}	Low	−0.412	−0.571[b]
	Mid	−0.309	−0.530[b]
	High	−0.264	−0.536[b]
IACC	Low	−0.452	−0.554[b]
	Mid	−0.183	−0.822[a]
	High	−0.086	−0.837[a]
	f_c	−0.575[b]	−0.389

[a]Correlation is significant at the 0.01 level (two-tailed).
[b]Correlation is significant at the 0.05 level (two-tailed).

Also evident from the scaling solution shown in Fig. 6 is that the modeled BRIRs (IDs 4–6) designed to approximate BRIRs measured from a particular room do not exactly match the percepts elicited by the measured BRIRs (IDs 1–3). In general, the modeled BRIRs lie slightly higher on Dimension 1 and slightly lower on Dimension 2, which based on the proposed interpretation of the two dimensions of the perceptual scale suggests that the modeled BRIRs are perceived as having slightly longer reverberation time and slightly less diffusivity than the measured BRIRs. For Dimension 1, this relationship is consistent with the observed physical values of T_{60} in the high frequencies, where modeled BRIRs had approximately 23% greater reverberation time than measured BRIRs (see Table II). For Dimension 2, the relationship is consistent with the observed physical values of IACC in the mid- and high frequencies, where modeled BRIRs had approximately 16% and 9% greater IACC values (see Table II). Although these perceptual mismatches between measured and modeled rooms do suggest shortcomings of the modeling procedures, it is important to note that the mismatches are limited to two perceptual dimensions with relatively clear physical correlates. As such, methods to improve the perceptual match should be relatively straightforward.

Of further note is the close proximity of modeled BRIRs using individualized HRTFs (IDs 1 and 4) to those using nonindividualized HRTFs (IDs 2, 3, 5, and 6) in Fig. 6. This suggests that the perceptual effects of manipulating properties of the reverberant sound are much larger than the effects due to the potentially degraded spatial rendering of the direct-path and early reflections with nonindividualized HRTFs. This result has important implications for the implementation of virtual room simulations, suggesting that individualized HRTFs may not be required to effectively simulate different room-acoustic environments when the source direction is fixed.

An additional and potentially interesting aspect of the perceptual scale shown in Fig. 6 is the relative location of room 14: a room with no late reverberant energy (see Table I for details). This "degraded" room is, in fact, closer to the measured rooms (1–3) than any of the modeled rooms (4–6) expressly designed to emulate the measured rooms. Although the enhanced match on Dimension 1 is perhaps predictable, based on the interpretation that Dimension 1 is a direct perceptual correlate of reverberation time and that room 14's

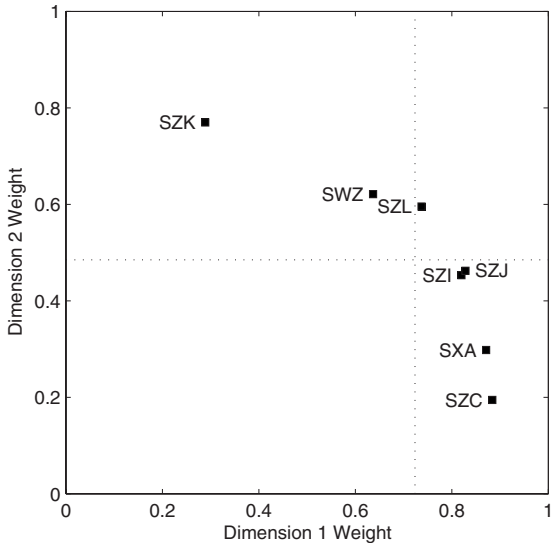


FIG. 7. Individual listener weights for the two-dimensional scaling solution (R^2=0.830). Dotted lines indicated the mean weight for each dimension.

reverberation time is reduced given the truncation of its late reverberant energy, it also clearly suggests that simulation of late reverberant energy may be relatively unimportant for the overall accurate perceptual recreation of small-room acoustics.

The relative importance of the two dimensions in the scaling solution for each listener are shown in Fig. 7. Substantial differences in the importance, or weight, each listener placed on the dimensions of the scaling solution are evident. The majority of listeners tended to place greater weight on Dimension 1, and relatively less weight on Dimension 2. The mean weight across all listeners for Dimension 1 was 0.72 (SD=0.21). For Dimension 2, the mean weight was 0.49 (SD=0.20). This suggests that for most listeners, aspects of room reverberation time are perhaps the primary bases for judging the similarities between room simulations in this study. Some listeners, however, appear to place approximately equal weight on the two dimensions, and one listener (e.g., SZK) appears to base their judgments of room similarity primarily on the perhaps spatially oriented aspects of Dimension 2.

IV. DISCUSSION

A. Physical testing

1. BRIR comparisons

Although the simplified room modeling techniques described in this study were capable of producing reasonable approximations to measured BRIRs in a real-room, closer analysis of BRIR similarity revealed that the approximations were least good in the frequency extremes of the late reverberant energy (see Fig. 3). Differences in the high frequencies were most likely due to the method of modeling the late reverberant energy, which was limited to frequencies below 4 kHz. Explanations of the differences at low frequencies are more complicated, however. Some of the differences likely resulted from the mere fact that correlational procedures were used to assess similarity of signals that were designed to have uncorrelated late reverberant energy. Other explanations of the differences include potential errors in the estimation of absorption coefficients used to model the late reverberant energy, as well as the lack of diffraction effects in the model. Previous studies have demonstrated the sensitivity of room modeling techniques to accurate estimates of surface absorption properties as well as low-frequency errors due to the lack of diffraction modeling (Vorländer, 1995; Bork, 2000) Although precise explanations of these differences observed in the BRIRs at the frequency extremes is beyond the scope of this study, it is also important to recognize the generally high degree of similarity observed using the same correlational procedures between measured and modeled BRIRs in the midfrequencies (400–6000 Hz). This result is encouraging, because it suggests that even greatly simplified room modeling techniques can produce reasonable physical matches to measured BRIRs from simple room environments at least over a somewhat restricted mid-frequency region.

2. Acoustical parameter comparisons

Given that direct comparison of the BRIRs demonstrated differences between measured and modeled results, it is not surprising that room-acoustic parameters derived from the BRIRs also show differences between models and measurements, as displayed in Table II. To better interpret the magnitude of these differences, they were compared relative to psychophysically determined just-noticeable differences (JNDs) for each parameter in isolation (see Table III). Two recent evaluative comparisons of commercially available room modeling software have used a similar approach (Vorländer, 1995; Bork, 2000). From the results shown in Table III, it is apparent that only the differences in C_{50} and IACC parameters were greater than 1 JND. This suggests that these rooms would likely have been discriminable based on either of these parameters in isolation, and not any of the other parameters (e.g. T_{60}, T_c, and f_c). Of course comparison of particularly the C_{50}, T_{60}, and T_c parameters from this study that were estimated from the response of a directional microphone (i.e., placed in the ear) to JND values from other studies where this was not the case (Seraphim, 1958; Cox *et al.*, 1993; Bradley *et al.*, 1999) must be made with considerable caution. The IACC comparison is perhaps more valid, given that comparable source directionality (at least in the low frequencies) and binaural recording techniques were used both to estimate the parameters in this study and by Okano (2002) to estimate IACC JND. The observed difference in IACC between measured and modeled rooms is most likely related to the directional properties of the measurement sound source at high frequencies (see Fig. 1). Finally, it is important to note that the differences between measured and modeled rooms in this study, expressed in terms of parameter JNDs, are within the range of differences reported in recent evaluative comparisons of various commercially available auralization packages relative to parameters based on measurements from real auditoria (Vorländer, 1995; Bork, 2000).

3. Independence of acoustical parameters

Although a large number of room-acoustic parameters were estimated from BRIR measurements in this study (see Fig. 4), results from the principal components analysis suggest that there are really only four independent parameters for this set of rooms: broadband T_{60}, broadband C_{50},f_c, and high-frequency IACC. Given that the sample of rooms in this study was in no way intended to be representative of all moderate-sized rooms, it is premature to extend this result to other sets of listening rooms. Nevertheless, there are potentially important similarities between this set of independent parameters, and results from four other studies based on physical measurements from mostly concert hall listening environments (Schroeder *et al.*, 1974; Ando and Schroeder, 1985; Beranek, 2004; Cerdá *et al.*, 2009): All find a relatively small number of independent parameters (2–6) and all sets of independent parameters include some measure of reverberation time and some measure of interaural coherence. These results are potentially important because they may suggest that a small set of independent physical parameters are invariant across a wide range of listening environments, and thus may be of considerable benefit to understanding general perceptual aspects of room acoustics. Beyond this, it is difficult to find other similarities in the results across studies, although Ando and Schroeder (1985), Beranek (2004), and Cerdá *et al.* (2009) did all report that strength factor was an additional independent physical parameter. Strength factor was not considered in Schroeder *et al.*, 1974, or in the current study, because in both studies sound level was equalized at the ear across all stimuli during psychophysical testing, thus removing strength factor as a basis for similarity judgments.

B. Psychophysical testing

1. Perceptual scale results

Even though the simplified methods used to model room acoustics did not produce exact physical matches to measured BRIRs, the modeling techniques still provide valuable insight into the perceptual differences between rooms—either simulated or real—with different acoustical properties. Based on the results of the MDS analysis of the participants' disimilarity ratings of all possible pairs of rooms, it is concluded that the scale of perceptual differences between the rooms is two-dimensional. Dimension 1 of this perceptual scale appears to be related to reverberation time and Dimension 2 appears to be related to sound spaciousness within the room. Overall, these results are consistent with several aspects of previous work in which perceptual scales of mostly concert hall listening environments have been estimated. First, the perceptual scales of room acoustics appear to be relatively low-dimensional. That is, a relatively few number of independent perceptual aspects of the room's acoustical properties comprise the entire room-acoustic percept. In complex listening environments, such as concert halls, there appears to be a greater number of independent perceptual parameters, although the exact number and makeup of the set of parameters has been an active area of study over the past quarter-century or more (Beranek, 1992; Cerdá *et al.*, 2009). In one study that examined perceptual scales of the acoustics of smaller listening rooms, the scale was found to be two-dimensional (Berkley and Allen, 1993). A second area of consensuses across nearly all studies is the primacy of reverberation time in percepts related to room acoustics. In this study, reverberation time appears to account for the majority of variance in most listeners' judgments of acoustical similarity. Work by Berkley and Allen (1993) also in small rooms demonstrated a similarly strong relationship between reverberation time and perceptual similarity. A number of other studies of concert hall acoustics draw similar conclusions regarding the perceptual importance of reverberation time (see Beranek, 1992 for review), although debate as to the exact perceptual relationship between reverberation time and other room-acoustic parameters continues. In general, this apparent primacy of reverberation time in the perceptual aspects of room acoustics is clear testament to practical importance placed on the physical quantification of reverberation time in a variety of listening environments dating at least to the work of Sabine (Sabine, 1922).

At least two aspects of the derived perceptual scale in this study depart from results of previous studies related to perceived small-room acoustics. First, the interpretation of Dimension 2 as a spatial dimension of perceived room acoustics is inconsistent with the second dimension of the perceptual scale reported by Berkley and Allen (1993). Although both studies conclude that the primary determinant of perceived small-room acoustics is reverberation time, Berkley and Allen (1993) suggested that a secondary determinant is related to a spectral variance parameter, which has been shown by other to relate to the ratio of direct-to-reverberant sound energy (Jetzt, 1979), instead of the spatial aspects of Dimension 2 reported here. One likely explanation for this difference is that the Berkley and Allen (1993) scaling study was conducted under monaural listening conditions, which would have eliminated any binaural information that is critically important for reproduction of the spatial aspects of room acoustics. Had spatial information been eliminated in the current study through monaural sound presentation, closer agreement on Dimension 2 of the scaling solutions in the two studies may have been observed. It is also interesting to note that the measured and modeled BRIRs in this study differed most along Dimension 2, suggesting that these two classes of BRIRs are perhaps most different perceptually in terms of spatial attributes related to IACC. A second departure from the results of past work is the seemingly minimal contribution of spectral/timbre aspects to perceived small-room acoustics. This result is surprising, given that results of past work suggest that spectral coloration caused by the acoustics of small rooms is an important perceptual aspect of such rooms (Olive and Toole, 1989; Bech, 1995, 1996). Although results from the physical parameter analyses in this study do suggest that timbre (as measured by f_c) is an independent physical parameter in this set of rooms, it does not appear to be an independent perceptual parameter. Instead, f_c is found to be related to Dimension 1 of the scaling solution, but more weakly related than reverberation time. As such, one might conclude that timbre does, in fact, contribute to the perceptual aspects of small-room acoustics in the current

study, but it does so much less than reverberation time and is not perceptually independent of reverberation time, at least for this set of room simulations. Recent data from Rumsey *et al.* (2005) show similar inter-relationships between spatial and timbral aspects of reproduced sound, although these authors find that timbre tended to dominate space under conditions of degraded multi-channel audio reproduction.

An additional important result from the perceptual scaling analysis in this study is that individual listeners differ considerably in the relative importance they place on the two perceptual dimensions related two reverberation time and spaciousness. Although most listeners weight the reverberation time dimension most heavily, clear exceptions to this rule were observed, where some listeners placed increased weight on the spaciousness dimension. Because past studies of preference in room acoustics have not generally used the multidimensional methods to analyze individual differences implemented in this study, results are difficult to compare directly. Reports of large individual differences in concert hall acoustics preference appear to be relatively common (Wilkens, 1977; Barron, 1988; Morimoto *et al.*, 1988), however.

Finally, it is important to note that the perceptual scaling results reported here were determined from a single source/listener location within each room, using a single set of speech source material. As a result, the sensitivity of these particular scaling results to source location or source material is not currently known, although past studies of concert hall preference have demonstrated effects of both factors (Yamaguchi, 1972). Clearly this is an area in need of study within smaller listening environments.

2. Practical implications

Results from the psychophysical scaling analysis have a number of practical implications for virtual listening simulation of small rooms. First, the greatly simplified room modeling techniques used here still provided reasonable matches to measured rooms. Physically, the matches in room-acoustic parameters were generally similar to the matches typically seen from other room modeling software packages (Vorländer, 1995; Bork, 2000). Perceptually, the modeled rooms did differ from the measured rooms, but only along the two dimensions of the perceptual similarity space thought to relate to reverberation time and spaciousness. Had very poor perceptual matches between measured and modeled rooms been present, one might have expected that the scaling solution would have included an additional dimension that independently differentiated between measured and modeled rooms: a "realism" dimension. Such a dimension was not observed. This result is encouraging, because it suggests that optimization of the model to perceptually match the measured response characteristics should be easy to accomplish via appropriate adjustments to the physical acoustic parameters of T_{60} and IACC that closely correspond to the two perceptual dimensions. Evaluation of such perceptual optimization strategies applied to room-acoustic modeling techniques, perhaps in conjunction with comparisons to results from more physically sophisticated models, is an area for additional study.

A second implication from these room scaling results is that individualized HRTFs do not appear to be necessary for realistic room simulation, since both measured and modeled BRIRs with nonindividualized HRTFs fall very near those with individualized HRTFs in the scaling solution (Fig. 6). Given the logistical challenges of obtaining individualized HRTF sets for each potential user in virtual listening simulations, this result is of particular practical importance. It is essential to note, however, that the testing situation in this study involved only a single static source location directly in front of the listener. As such, the relevance of the likely small directional errors in source rendering resulting from the use of nonindividualized HRTF did not play a major role in listeners' judgments of room similarity. It is also likely that in listening situations such as those reproduced here, the directional information contained in at least some of the early reflections may be suppressed through a process commonly known as the precedence effect (Wallach *et al.*, 1949; Litovsky *et al.*, 1999), thus further minimizing the need for highly accurate simulation of the directional information in the early reflections using individualized HRTFs. Although there can be no doubt that a variety of room simulation applications would be well-served through accurate simulation of a single fixed-direction sound source, other more complicated applications with variable source direction would likely benefit from the increased directional accuracy afforded by individualized HRTFs, as has been demonstrated in previous work within anechoic space (Wenzel *et al.*, 1993).

A final somewhat more tenuous implication relates to the simulation of late reverberant energy. In the physical analyses of the measured and modeled BRIRs, it was noted that one significant source of error appeared to stem from the modeling of the late reverberant energy. Hence, modeling techniques for this portion of the response are an obvious place for improvement. Surprisingly, however, the modeled room stimulus which had its late reverberant energy artificially removed (room 14) was closer on the perceptual scale than other modeled rooms which included simulated late reverberation. Perhaps no reverberation, which is analogous to the simulation situations originally described by Allen and Berkley (1979), is better than poor reverberation. Clearly additional research is required to more fully examine this effect.

V. CONCLUSIONS

The simple room simulation methods described in this study provide a reasonable physical approximation to BRIRs measured in a real room. Modeling errors generally increased at the frequency extremes. When room-acoustic parameters were estimated from the BRIRs, differences between modeled and measured rooms for most parameters were on the order of those reported for other room modeling algorithms (Bork, 2000). Potential causes for these differences are not fully understood, but likely include model limitations related to directional sound sources, as well as inadequacies in the model's treatment of late reverberant energy.

Four independent acoustics parameters derived from the measured or modeled BRIRs resulting from the 15 room simulations examined in this study were found. These parameters were broadband T_{60}, broadband C_{50},f_c, and high-frequency IACC.

MDS results suggest that only two dimensions are perceptually relevant for judging the similarity between the 15 room simulations examined in this study. Dimension 1 of the scaling solutions was highly correlated with T_{60} ($|r|>0.94$). Dimension 2 was highly correlated with mid- and high-frequency IACC ($|r|>0.82$). Measured and modeled rooms were relatively close together in the scaling solution, suggesting a good degree of perceptual similarity.

Most listeners based their judgments of room similarity primarily on reverberation time (Dimension 1), although relatively large individual differences were observed.

Effects of spatial rendering quality (individualized HRTFs) were small, which has important practical implications for virtual auditory display and room auralization applications.

ACKNOWLEDGMENTS

Thanks to Jen Junion-Dienger for her assistance in data collection, as well as to 3 anonymous reviewers for their comments on an earlier version of this manuscript. Work supported by NIH-NIDCD (R03 DC005709 and R01 DC008168).

Allen, J. B., and Berkley, D. A. (**1979**). "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950.

Ando, Y., and Schroeder, M. R. (**1985**). *Concert Hall Acoustics* (Springer-Verlag, Berlin).

ANSI-S1.11 (**2004**). "Specification for octave-band and fractional-octave-band analog and digital filters" (American National Standards Institute, New York).

ANSI-S3.9 (**1989**). "American National Standard specification for audiometers" (American National Standards Institute, New York).

Barron, M. (**1988**). "Subjective study of British symphony concert halls," Acustica **66**, 1–14.

Bech, S. (**1995**). "Timbral aspects of reproduced sound in small rooms I," J. Acoust. Soc. Am. **97**, 1717–1726.

Bech, S. (**1996**). "Timbral aspects of reproduced sound in small rooms II," J. Acoust. Soc. Am. **99**, 3539–3549.

Beranek, L. L. (**1992**). "Concert hall acoustics—1992," J. Acoust. Soc. Am. **92**, 1–40.

Beranek, L. L. (**2004**). *Concert Halls and Opera Houses: Music, Acoustics, and Architecture* (Springer, New York).

Berkley, D. A., and Allen, J. B. (**1993**). "Normal listening in typical rooms: The physical and psychophysical correlates of reverberation," in *Acoustical Factors Affecting Hearing aid Performance*, edited by G. A. Studebaker and I. Hockberg (Allyn and Bacon, Boston), pp. 3–14.

Bilger, R. C., and Wang, M. D. (**1976**). "Consonant confusions in patients with sensorineural hearing loss," J. Speech Hear. Res. **19**, 718–748.

Bork, I. (**2000**). "A comparison of room simulation software—The 2nd round robin on room acoustical computer simulation," Acta Acust. **86**, 943–956.

Bradley, J. S., Reich, R., and Norcross, S. G. (**1999**). "A just noticeable difference in C_{50} for speech," Appl. Acoust. **58**, 99–108.

Carroll, J. D., and Chang, J.-J. (**1970**). "Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart–Young" decomposition," Psychometrika **35**, 283–319.

Cerdá, S., Giménez, A., Romero, J., Cibrián, R., and Mirallese, J. L. (**2009**). "Room acoustical parameters: A factor analysis approach," Appl. Acoust. **70**, 97–109.

Cox, T. J., Davies, W. J., and Lam, Y. W. (**1993**). "The sensitivity of listeners to early sound field changes in auditoria," Acustica **79**, 27–41.

Emiroglu, S., and Kollmeier, B. (**2008**). "Timbre discrimination in normal-hearing and hearing-impaired listeners under different noise conditions," Brain Res. **1220**, 199–207.

Grey, J. M. (**1977**). "Multidimensional perceptual scaling of musical timbres," J. Acoust. Soc. Am. **61**, 1270–1277.

Grey, J. M., and Gordon, J. W. (**1978**). "Perceptual effects of spectral modifications on musical timbres," J. Acoust. Soc. Am. **63**, 1493–1500.

Hammershøi, D., and Møller, H. (**1996**). "Sound transmission to and within the human ear canal," J. Acoust. Soc. Am. **100**, 408–427.

Hartmann, W. M. (**1983**). "Localization of sound in rooms," J. Acoust. Soc. Am. **74**, 1380–1391.

Hartmann, W. M., and Wittenberg, A. (**1996**). "On the externalization of sound images," J. Acoust. Soc. Am. **99**, 3678–3688.

Heinz, R. (**1993**). "Binaural room simulation based on an image source model with addition of statistical methods to include the diffuse sound scattering of walls and to predict the reverberant tail," Appl. Acoust. **28**, 145–159.

ISO-3382 (**1997**). "Acoustics—Measurement of the reverberation time of rooms with reference to other acoustical parameters" (International Organization for Standardization, Geneva).

Jetzt, J. J. (**1979**). "Critical distance measurement of rooms from the sound energy spectral response," J. Acoust. Soc. Am. **65**, 1204–1211.

Kempster, G. B., Kistler, D. J., and Hillenbrand, J. (**1991**). "Multidimensional scaling analysis of dysphonia in two speaker groups," J. Speech Hear. Res. **34**, 534–543.

Kewley-Port, D., and Atal, B. S. (**1989**). "Perceptual differences between vowels located in a limited phonetic space," J. Acoust. Soc. Am. **85**, 1726–1740.

Kleiner, M., Dalenbäck, B., and Svensson, P. (**1993**). "Auralization—An overview," J. Audio Eng. Soc. **41**, 861–875.

Kulkarni, A., and Colburn, H. S. (**1998**). "Role of spectral detail in sound-source localization," Nature (London) **396**, 747–749.

Kuttruff, H. (**2000**). *Room Acoustics* (Spon, London).

Lam, Y. W. (**2005**). "Issues for computer modeling of room acoustics in non-concert hall settings," Acoust. Sci. & Tech. **26**, 145–155.

Langendijk, E. H., and Bronkhorst, A. W. (**2000**). "Fidelity of three-dimensional-sound reproduction using a virtual auditory display," J. Acoust. Soc. Am. **107**, 528–537.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.

Møller, H. (**1992**). "Fundementals of binaural technology," Appl. Acoust. **36**, 171–218.

Møller, H., Hammershøi, D., Jensen, C. B., and Sørensen, M. F. (**1995**). "Transfer characteristics of headphones measured on human ears," J. Audio Eng. Soc. **43**, 203–217.

Morimoto, M., Maekawa, Z.-i., Tachibana, H., Yamasaki, Y., Hirasawa, Y., and Pösselt, C. (**1988**). "Preference test of seven European concert halls," J. Acoust. Soc. Am. **84**, S129.

Moulder, R. (**1991**). "Sound-absorptive materials," in *Handbook of Acoustical Measurements and Noise Control*, edited by C. M. Harris (McGraw-Hill, New York), pp. 30.31–31.31.

Nabelek, A. K., and Robinson, P. K. (**1982**). "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am. **71**, 1242–1248.

Okano, T. (**2002**). "Judgments of noticeable differences in sound fields of concert halls caused by intensity variations in early reflections," J. Acoust. Soc. Am. **111**, 217–229.

Olive, S. E., and Toole, F. E. (**1989**). "The detection of reflections in typical rooms," J. Audio Eng. Soc. **37**, 539–553.

Peutz, V. M. A. (**1971**). "Articulation loss of consonants as a criterion for speech transmission in a room," J. Audio Eng. Soc. **19**, 915–919.

Rife, D. D., and Vanderkooy, J. (**1989**). "Transfer-function measurement with maximum-length sequences," J. Audio Eng. Soc. **37**, 419–444.

Rindel, J. H. (**2000**). "The use of computer modeling in room acoustics," J. Vibroeng. **3**, 219–224.

Rumsey, F., Zielinski, S., Kassier, R., and Bech, S. (**2005**). "On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality," J. Acoust. Soc. Am. **118**, 968–976.

Sabine, W. C. (**1922**). "Reverberation," *Collected Papers on Acoustics* (Harvard University Press, Cambridge, MA).

Schroeder, M. R. (**1970**). "Synthesis of low-peak-factor signals and binary sequences with low autocorrelation," IEEE Trans. Inf. Theory **16**, 85–89.

Schroeder, M. R., Gottlob, D., and Siebrasse, K. F. (**1974**). "Comparative study of European concert halls: correlation of subjective preference with

geometric and acoustic parameters," J. Acoust. Soc. Am. **56**, 1195–1201.
Seraphim, H. P. (**1958**). "Untersuchungen über die Unterschiedsschwelle exponentiellen Abklingens von Rauschbandimpulsen (Investigations on the difference limen of exponentially decaying bandlimited noise pulses)," Acustica **8**, 280–284.
Soli, S. D., and Arabie, P. (**1979**). "Auditory versus phonetic accounts of observed confusions between consonant phonemes," J. Acoust. Soc. Am. **66**, 46–59.
Vorländer, M. (**1995**). "International round robin on room acoustical computer simulations," in 15th International Congress on Acoustics, Trondheim, Norway.
Vorländer, M. (**2008**). *Auralization* (Springer-Verlag, Berlin).
Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**). "The precedence effect in sound localization," Am. J. Psychol. **62**, 315–336.
Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**, 111–123.
Wightman, F. L., and Kistler, D. J. (**1989**). "Headphone simulation of free-field listening: I. Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.
Wilkens, H. (**1977**). "Mehrdimensionale beschreibung subjektiver beurteilungen der akustik von konzertsälen (Multidimensional description of subjective judgments of the acoustics of concert halls)," Acustica **38**, 10–23.
Yamaguchi, K. (**1972**). "Multivariate analysis of subjective and physical measures of hall acoustics," J. Acoust. Soc. Am. **52**, 1271–1279.
Zahorik, P. (**2002**). "Assessing auditory distance perception using virtual acoustics," J. Acoust. Soc. Am. **111**, 1832–1846.
Zahorik, P. A., Wightman, F. L., and Kistler, D. J. (**1995**). "On the discriminability of virtual and real sound sources," in *Proceedings of the ASSP (IEEE) Workshop on Application of Signal Processing to Audio and Acoustics* (IEEE, New York).

Effects of envelope bandwidth on the intelligibility of sine- and noise-vocoded speech

Pamela Souza
Department of Speech and Hearing Sciences, University of Washington, 1417 NE 42nd Street, Seattle, Washington 98105

Stuart Rosen
Speech, Hearing and Phonetic Sciences, Division of Psychology and Language Sciences, UCL, 2 Wakefield Street, London WC1N 2PF, United Kingdom

(Received 19 June 2008; revised 16 May 2009; accepted 26 May 2009; corrected 16 December 2009)

The choice of processing parameters for vocoded signals may have an important effect on the availability of various auditory features. Experiment 1 varied envelope cutoff frequency (30 and 300 Hz), carrier type (sine and noise), and number of bands (2–5) for vocoded speech presented to normal-hearing listeners. Performance was better with a high cutoff for sine-vocoding, with no effect of cutoff for noise-vocoding. With a low cutoff, performance was better for noise-vocoding than for sine-vocoding. With a high cutoff, performance was better for sine-vocoding. Experiment 2 measured perceptibility of cues to voice pitch variations. A noise carrier combined with a high cutoff allowed intonation to be perceived to some degree but performance was best in high-cutoff sine conditions. A low cutoff led to poorest performance, regardless of carrier. Experiment 3 tested the relative contributions of comodulation across bands and spectral density to improved performance with a sine carrier and high cutoff. Comodulation across bands had no effect so it appears that sidebands providing a denser spectrum improved performance. These results indicate that carrier type in combination with envelope cutoff can alter the available cues in vocoded speech, factors which must be considered in interpreting results with vocoded signals.
 [DOI: 10.1121/1.3158835]

PACS number(s): 43.66.Lj, 43.71.Es, 43.71.Bp [RLF] Pages: 792–805

I. INTRODUCTION

Noise- and tone-vocoded signals are used to study the use of temporal cues in normal-hearing (e.g., Shannon *et al.*, 1995; Souza and Turner, 1998; Gallun and Souza, 2008) and hearing-impaired listeners (e.g., Turner *et al.*, 1995; Souza and Boike, 2006) as well as to simulate the information available to cochlear implant users (e.g., Rosen *et al.*, 1999; Shannon *et al.*, 2004; Green *et al.*, 2007). To create these signals, an envelope is extracted from each band of a filter bank and used to modulate a carrier that is either a sine wave at the center frequency of the band or a white noise subsequently filtered to the channel bandwidth. There has been much recent interest in delineating the various factors that are important in accounting for differences between sine- and noise-vocoded speech, and also some areas of disagreement between previous studies. Dorman *et al.*, 1997 is often cited as demonstrating no effect of carrier type for consonants, vowels, or sentences. In contrast, recent studies have found better sentence and/or vowel recognition with sine than with noise carriers (Gonzalez and Oliver, 2005; Chang and Fu, 2006; Whitmal *et al.*, 2007; Stone *et al.*, 2008). For consonants, Whitmal *et al.* (2007) found that sine carriers provided better performance in noise but equivalent performance in quiet. Under some circumstances then, sine-vocoded and noise-vocoded speech do not seem to provide the same information, at least for normal-hearing listeners.

To understand these differences, we briefly consider the acoustic form of these signals for a relatively small number of channels, say, 6 or less, where we expect little or no resolution of the harmonics of voiced speech by the analysis filter bank (Fig. 1). Of crucial importance, and interacting with the carrier type, is the cutoff frequency of the envelope smoothing filter that typically follows rectification, as this determines the range of modulation frequencies available in the envelope signal.

A. Vocoding with a sine carrier

When the envelope cutoff is high in comparison to the talker's fundamental frequency (F0), voiced speech will result in envelopes containing amplitude fluctuations corresponding in rate to the voice pitch of the talker. When such an envelope is multiplied against the sinusoidal carrier, sidebands consisting of the sum and difference frequencies of each spectral component in the envelope spectrum and the carrier frequency will be created (as well as a strong component at the carrier frequency due to the strong dc component in the envelope spectrum). Thus the spectrum of each resulting channel output will consist of a series of harmonic-like spectral components centered at the carrier frequency, plus and minus the fundamental frequency. Different from what happens in natural speech, the spectral components squeeze together and expand around their central frequency rather than sweeping all in the same direction during the characteristic F0 glides that constitute intonation. Informal observa-

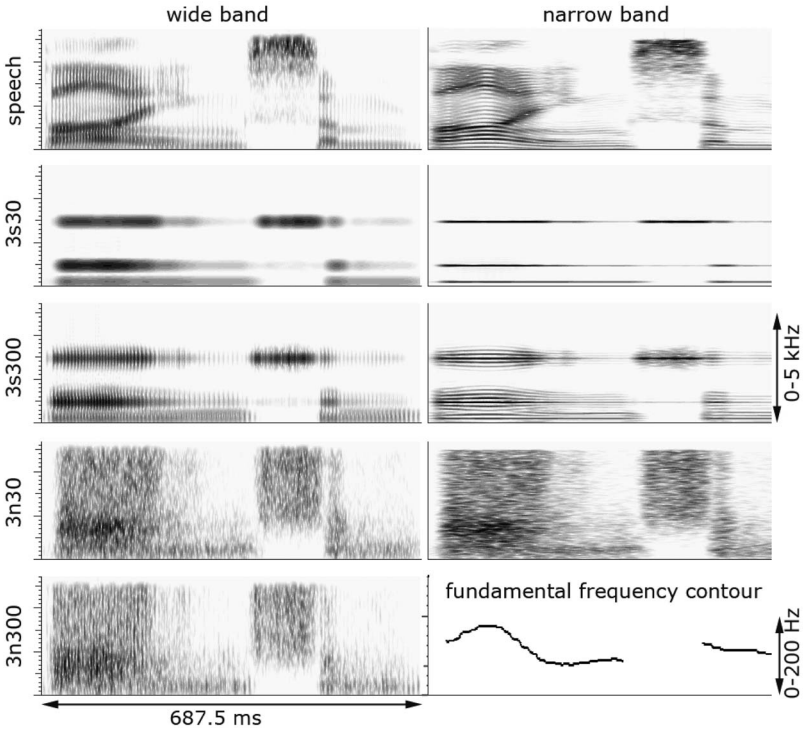


FIG. 1. Wide- and narrow-band spectrograms of various versions of the phrase "buying some" extracted from a male speaker uttering the sentence "They're buying some bread." The top row shows the original speech, followed by examples of the processing used in experiment 1, all based on three-channel sine or noise vocoders with two different envelope smoothing cut-off frequencies (30 and 300 Hz). Hence condition 3s300 refers to a three-channel sine vocoder with an envelope cutoff frequency of 300 Hz. The 30 Hz cutoff was chosen to be significantly below the fundamental frequency (F0) of the talker, and the 300 Hz cutoff chosen to be significantly above. The F0 contour of the utterance can be seen at lower right (narrow-band spectrograms of 3n30 and 3n300 stimuli are visually indistinguishable).

tions suggest that such signals lead to a percept that has a reasonably strong pitch, making information about voicing and intonation accessible to a listener.

Quite apart from the explicitly phonetic information, the common modulation sine components receive during voiced speech may also have a beneficial impact on intelligibility. Carrell and Opie (1992) showed that such a common modulation, at least at low enough modulation rates, can improve the intelligibility of sine-wave speech, presumably by making the components cohere better (but note the important acoustic differences between sine-vocoded speech and sine-wave speech).

For aperiodic speech sounds, the range of fluctuation rates in the envelope signal will depend in a straightforward way on the envelope filter cutoff. Higher cutoffs will lead to faster fluctuations, but these will be random in nature. On the one hand, these higher-rate envelope fluctuations may obscure the slower envelope fluctuations. On the other hand, higher-rate envelope fluctuations may cue the presence of aperiodic energy more effectively, a strong cue to voicelessness and fricative manner. However, it is likely that the effect of envelope fluctuations, either for or against aperiodic energy, operates only in the absence of other spectral shape cues (i.e., only in single-channel vocoding) because voiceless excitation tends to have much more energy in high frequency regions than low, the opposite trend to that observed for voiced sounds. Indeed, there is good evidence for the utility of envelope fluctuations above 20 Hz when using a single channel, but not when using higher numbers of channels (Shannon *et al.*, 1995; Fu and Shannon, 2000).

The situation is much simpler when the envelope filter cutoff is low in comparison to F0. Here, any within-channel information about speech periodicity in the envelopes arising from voiced speech will be eliminated and only slow fluctuations will be transmitted. This will result in the output spectrum being dominated by what is essentially a small number of sinusoids varying slowly in amplitude, one for each channel in the vocoder.

B. Vocoding with a noise carrier

When noise is used as a carrier, inherent fluctuations in the noise will be superimposed on the envelope. It seems plausible that these fluctuations (non-existent for a sine carrier) could interfere with or obscure some envelope cues. On the other hand, noise-vocoded consonants result in similar auditory nerve responses to natural speech consonants (Loebach and Wickesberg, 2006) so perhaps the carrier fluctuations are irrelevant.

Varying the envelope filter bandwidth again has its major effects for voiced speech. For envelope filter cutoffs low in comparison to F0, there should be *no* direct cues for periodicity, hence little or no percept of voice pitch. When envelope filter cutoffs are high in comparison to F0, cues to periodicity and intonation are signaled through amplitude modulations in the noise carrier which may not be very deep, and even in the best situations lead to relatively weak pitches (Burns and Viemeister, 1976, 1981; Patterson *et al.*, 1978). It therefore seems likely that the voice pitch of the talker will be much less salient in noise-vocoded than in sine-vocoded speech at least for high envelope cutoffs. Supporting evidence for this comes from work by Stone *et al.* (2008) who investigated sentence recognition in a single-competing-talker background for six-channel noise and tone carriers. Although this study does not resolve the issue of higher-frequency envelope contributions to speech recognition *per se*, it would be expected that F0 cues would aid talker separation and thus improve performance. In fact, varying envelope bandwidth had much larger effects for sine than noise carriers.

TABLE I. Center and lower-to-upper cutoff frequencies (in hertz) used for the vocoder processing. For each condition the lowest cutoff frequency was 100 Hz and the highest cutoff frequency was 5000 Hz.

	Center frequency (Hz)				
Band	1	2	3	4	5
2	392	2294			
3	269	1005	2984		
4	219	643	1531	3399	
5	192	481	1005	1955	3673
	Lower-to-upper cutoff frequencies (Hz)				
Band	1	2	3	4	5
2	100–1005	1005–5000			
3	100–548	548–1755	1755–5000		
4	100–392	392–1005	1005–2294	2294–5000	
5	100–315	315–705	705–1410	1410–2687	2687–5000

It is also important to note that varying envelope bandwidth should have little impact on the spectrum of noise-vocoded speech. Although the modulation of the envelope by the noise carrier still leads to sidebands, multiplying a white noise by any signal still results in a white noise. Here, the spectral properties of each band are determined by the properties of the filters used to limit the bandwidth of each channel *after* modulation.

In summary, then, envelope cutoff frequency and carrier type, as well as number of bands, interact in a complex way and require further study, especially as regards their effects on intelligibility. Generally speaking, we expect much bigger effects of envelope bandwidth for sine-vocoded rather than noise-vocoded speech, because the bandwidth of the extracted envelopes is a crucial determinant of the temporal and spectral characteristics of sine-vocoded speech, but only affects the *temporal* properties of noise-vocoded speech.

Experiment 1 provides a direct assessment of these effects. Two follow-up experiments examine the contributions of fundamental-frequency variations (experiment 2) and carrier comodulation across bands (experiment 3) to the results found.

II. EXPERIMENT 1: ENVELOPE FREQUENCY CUTOFF VERSUS CARRIER TYPE

The purpose of experiment 1 was to vary envelope frequency cutoff, carrier type, and number of bands. We expected an interaction between carrier type and envelope cutoff. Previous work also suggested that such an interaction might depend on the type of speech materials.

A. Subjects

Subjects were 16 adults (12 females and 4 males) recruited from the student population at UCL. All were native speakers of southern British English. Subjects ranged in age from 19 to 52 years (mean 25 years). All but one listener had normal hearing, defined as pure-tone thresholds of 20 dB HL or better (see ANSI, 2006) at octave frequencies between 0.25 and 8 kHz. The exception was a single listener (aged 52 years) who had a pure-tone threshold of 30 dB HL at 8 kHz in one ear but who met the 20 dB HL inclusion criteria at all other test frequencies. None of the subjects had any prior experience with the test materials. All subjects were paid for their participation.

B. Stimuli and procedure

All test materials were spoken by the same female talker, who was a native speaker of southern British English. Stimuli were digitally recorded in a quiet room with sampling rates of 22.05 kHz for consonant and vowel stimuli and 48.1 kHz for sentences. The stimuli were vocoded using locally developed MATLAB software, as follows. Each file was digitally filtered into two, three, four, or five bands, using sixth-order Butterworth IIR filters. Filter spacing was based on equal basilar membrane distance (Greenwood, 1990) across a frequency range of 100–5000 Hz. Band center and cutoff frequencies are shown in Table I. Next, the output of each band was half-wave rectified and low-pass filtered (fourth-order Butterworth) at either 30 or 300 Hz to extract the amplitude envelope. The envelope was then multiplied by a carrier, either a tone at the band center frequency or a noise. The resulting signal (envelope $\times$ carrier) was filtered using the same bandpass filter as for the first filtering stage. The rms level was adjusted at the output of the filter to match the original analysis, and the signal was summed across bands. Sixteen different conditions were created, each with a unique combination of envelope cutoff frequency (30 or 300 Hz), carrier type (sine or noise), and number of bands (two, three, four, or five).

1. Consonant recognition

Consonant recognition was measured with a set of 20 syllables, including the consonants /b/, /tʃ/, /d/, /f/, /g/, /dʒ/, /k/, /l/, /m/, /n/, /p/, /r/, /s/, /ʃ/, /t/, /v/, /w/, /j/, /z/, and /ʒ/ in an /ɑCɑ/ context. These tokens were presented diotically through Sennheiser HD 25 SP headphones at a level between 65 and 70 dB SPL. On each trial, subjects heard a single syllable and selected among a set of orthographic choices using a computer mouse.

To familiarize them with the task, subjects began with a practice block, in which the 20 consonants were sampled from the various test conditions. Each test condition occurred at least once during the practice block. If the response to a trial was correct, visual correct answer feedback was provided and the processed consonant was played again, along with the unprocessed version. If the response was incorrect, the processed version was replayed and the subject was asked to try again. If the second response to that trial was incorrect, the correct answer was shown and the processed and unprocessed consonants were played.

In the test phase, the subject completed one block for each of the 16 test conditions. Each block contained 40 consonants (each token appeared twice). Each subject heard a different order of the conditions based on a Latin square. Correct answer feedback was not provided in the test phase.

2. Vowel recognition

Vowel recognition was measured with a set of ten vowels in a /bVd/ context: "bad" (/æ/), "bard" (/ɑː/), "bawd" (/ɔː/), "bead" (/iː/), "bed" (/e/), "bid" (/ɪ/), "bird" (/ɜː/), "bod" (/ɑ/), "booed" (/uː/), and "bud" (/ʌ/). Testing was similar to that described for consonants except the practice block sampled 30 /bVd/ words from the various test conditions and each test block contained 30 vowels (each token appeared three times).

3. Sentence recognition

Sentences were drawn from the ASL (MacLeod and Summerfield, 1990) and the BKB sentence lists (Bench and Bamford, 1979). Sentence testing was done in an open-set format. The subject repeated the sentence to the experimenter, who was seated in the same room and scored the responses using a computer program, which showed the three key words. The scoring screen was not visible to the subject. Listeners began with a practice block of ten sentences. After responding to a practice sentence, the subject heard the processed and unprocessed versions of that sentence. Each test block consisted of 15 ASL sentences and 16 BKB sentences, with each sentence played once without feedback. No list was repeated. Each sentence had three key words so the score for each condition was based on 93 words.

C. Results

For each reported analysis, Mauchly's (1940) test was evaluated and the Greenhouse–Geisser adjusted values were used if the assumption of sphericity was violated.

1. Consonant recognition

In Fig. 2, proportion correct for each condition is plotted as a function of number of bands. A repeated-measures analysis of variance comparing number of bands, carrier type, and envelope cutoff frequency showed significant main effects of number of bands ($F_{3,45}=275.16$, $p<0.005$), carrier type ($F_{1,15}=10.24$, $p=0.006$), and envelope cutoff frequency ($F_{1,15}=19.54$, $p<0.005$) with no significant three-way interaction ($F_{3,45}=1.60$, $p=0.202$). Carrier type

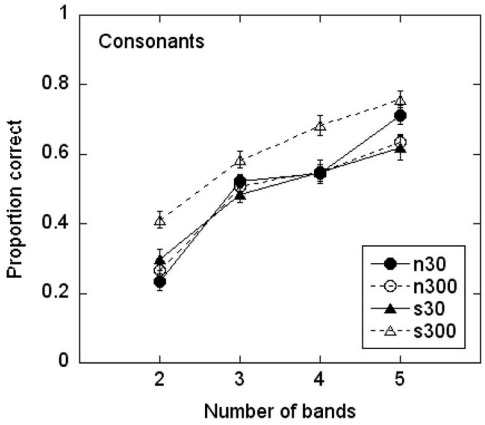


FIG. 2. Proportion correct for consonants as a function of number of bands. Circles show noise carriers; triangles show sine carriers. Filled symbols/solid lines show the 30 Hz envelope cutoff; open symbols/dashed lines show the 300 Hz envelope cutoff. Error bars show ±1 standard error.

interacted with envelope cutoff frequency ($F_{1,15}=23.74$, $p<0.005$) and with number of bands ($F_{3,45}=3.83$, $p=0.016$). Number of bands and envelope cutoff frequency did not interact ($F_{3,45}=1.14$, $p=0.341$).

To understand the nature of the interaction, separate two-way analysis of variance (ANOVAs) (bands × carrier type) were carried out for each envelope cutoff frequency. At a 30 Hz envelope cutoff frequency, performance improved with increasing band number ($F_{3,45}=160.54$, $p<0.005$) and there was no difference between a sine and noise carrier ($F_{1,15}=0.68$, $p=0.423$). A significant bands × carrier interaction [$F_{3,45}=3.84$, $p=0.016$] was caused by better performance ($t_{15}=2.23$, $=0.041$) for the five-band noise carrier, although this difference was no longer significant once Bonferroni-corrected. With a 300 Hz envelope cutoff frequency, performance improved with increasing band number ($F_{3,45}=110.40$, $p<0.005$) and was higher for the sine than for the noise carrier ($F_{1,15}=29.00$, $p<0.005$). The bands × carrier interaction was not significant ($F_{3,45}=1.13$, $p=0.348$).

Information transfer measures were also calculated and are shown in Fig. 3 for voicing (voiced and unvoiced), manner (stop, fricative, nasal, and glide), and place (front, middle, and back). For each condition, the analysis was performed on the composite confusion matrix representing data collapsed across 16 subjects. No place information was available until the signals contained more than two bands, consistent with the fact that place cues rely on spectral distinctions and spectro-temporal dynamics. The sine 300 condition was superior in most respects, particularly for voicing; even the two-band condition transmitted nearly 100% voicing information.

2. Vowel recognition

Results for vowels are shown in Fig. 4. These data were submitted to a repeated-measures analysis of variance (bands × carrier × envelope cutoff). The main effects of bands ($F_{3,39}=144.54$, $p<0.005$) and envelope cutoff frequency ($F_{1,13}=11.00$, $p=0.006$) were significant but the effect of carrier was not ($F_{1,13}=0.401$, $p=0.537$). The three-

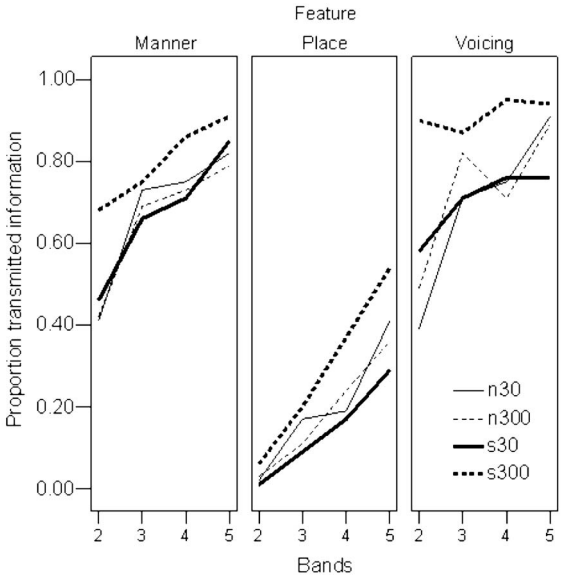


FIG. 3. Transmitted information for voicing, place, and manner for the consonant identification task.

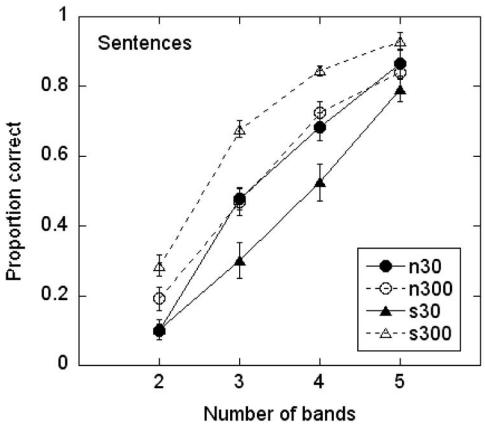


FIG. 5. Proportion correct for sentences as a function of number of bands. Circles show noise carriers; triangles show sine carriers. Filled symbols/solid lines show the 30 Hz envelope cutoff; open symbols/dashed lines show the 300 Hz envelope cutoff. Error bars show ±1 standard error.

way interaction was nonsignificant ($F_{3,39}=0.80$, $p=0.500$). Unlike the consonants, carrier type and cutoff frequency did not interact ($F_{1,13}=3.03$, $p=0.105$). Instead, the number of bands interacted with both envelope cutoff ($F_{3,39}=5.62$, $p=0.003$) and with carrier type ($F_{3,39}=3.61$, $p=0.021$).

Because bands interacted with both effects of interest, the comparisons of carrier type and envelope cutoff were obtained from two-way ANOVAs (carrier × envelope cutoff), one each for the two-band, three-band, four-band, and five-band stimuli. The carrier × envelope cutoff interaction was nonsignificant ($p>0.05$) in each case. Subjects performed better with the sine carrier than with the noise carrier only for the four-band stimulus ($p=0.001$). For each of the remaining band conditions, scores for the sine and noise carriers were statistically similar. Subjects performed better with the 300 Hz envelope frequency cutoff for four-band ($p=0.009$) and five-band ($p=0.002$) stimuli, but not for the two-band or three-band stimuli. This was partially determined by floor effects, as scores for the two-channel stimuli approached chance.

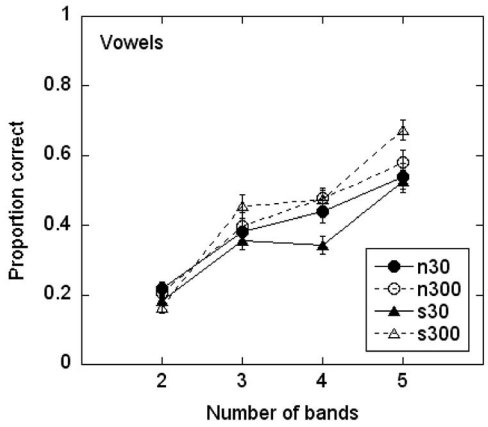


FIG. 4. Proportion correct for vowels as a function of number of bands. Circles show noise carriers; triangles show sine carriers. Filled symbols/solid lines show the 30 Hz envelope cutoff; open symbols/dashed lines show the 300 Hz envelope cutoff. Error bars show ±1 standard error.

We expected that vowels, which are identified by their spectrum, would be better identified as the number of bands increased. In that sense, the lack of improvement between three and four bands for the sine carrier was surprising. To ensure that this finding was not due to any artifact of the vowel token, three subjects repeated the testing with different exemplars from the female talker, as well as a different female talker. As for the original testing, there was little improvement between three and four bands for the sine-vocoded vowel stimuli. A close examination of Dorman *et al.*, 1997 shows a similar effect in which sine-vocoded natural vowels spoken by a single talker show no improvement between four and five bands. This probably occurred because the increase from three to four bands in the current study and from four to five bands in the study of Dorman *et al.* (1997) did not improve representation of the vowel spectra. To understand this, consider a simple example: the vowels /æ/ and /ɑ/ produced by a male talker from the Pacific Northwest. The first formant (F1) for both vowels is approximately 700 Hz, and the second formant (F2) is approximately 1200 Hz for /ɑ/ and 1700 Hz for /æ/ (Bor *et al.*, 2008). Referring to the band cutoff frequencies in Table I, these two vowels would probably not be distinguishable by a three-band vocoder, where they would produce a similar peak in band 2; nor by a four-band vocoder, where they would produce a similar peak in band 3. Only with the five-band vocoder, where F2 for /ɑ/ excites band 3, but F2 for /æ/ excites band 4, would we expect them to be distinguished on the basis of their vocoder response.

3. Sentence recognition

The differences among conditions are more pronounced for sentences (Fig. 5), compared to the vowels and consonants. These data were submitted to a repeated-measures analysis of variance (bands × carrier type × envelope cutoff). The main effects of bands ($F_{3,27}=246.83$, $p<0.005$) and envelope cutoff ($F_{1,9}=34.69$, $p<0.005$) were

significant but that of carrier type was not ($F_{1,9}$=0.91, p=0.366). The interaction between carrier type and envelope cutoff frequency ($F_{1,9}$=31.36, $p<0.005$) was significant. The interactions between bands and carrier type ($F_{3,27}$=1.08, p=0.373) and between bands and envelope cutoff frequency ($F_{3,27}$=2.76, p=0.062) were not significant. The three-way interaction was significant ($F_{3,27}$=4.74, p=0.009).

To analyze the various interactions, separate two-way ANOVAs (bands × carrier type) were completed for each envelope cutoff frequency. At a 30 Hz envelope cutoff frequency, scores improved with increasing band number ($F_{3,30}$=119.52, $p<0.005$) and were higher with a noise carrier ($F_{1,10}$=10.90, p=0.008). A significant bands × carrier interaction ($F_{3,30}$=4.46, p=0.010) was due to a floor effect that precluded any difference between two-band carriers (p=0.962), while scores were higher for the noise carrier than the tone carrier with three (p=0.015), four (p=0.002), and five bands (p=0.004). At a 300 Hz envelope cutoff frequency, scores improved with increasing band number ($F_{3,33}$=295.51, $p<0.005$) and were higher with a sine carrier ($F_{1,11}$=24.17, $p<0.005$). The bands × carrier interaction was not significant ($F_{3,33}$=1.71, p=0.184).

D. Discussion

Differences among conditions can be summarized as follows: For a 30 Hz envelope cutoff, performance was better for a noise carrier than a sine carrier; for a 300 Hz envelope cutoff, performance was better for a sine carrier than a noise carrier. We can also consider results for a single carrier as envelope cutoff frequency increases: This improves recognition for a sine carrier, but not for a noise.

To understand these effects, consider the differences between the various conditions. The first is that with the 300 Hz envelope cutoff in combination with a sine carrier, cues to voice fundamental frequency (voicing and intonation) should be available to the listener (a question we explicitly address in experiment 2). Intonation itself has been shown to make a small contribution to sentence intelligibility (Hillenbrand, 2003) but should not matter much for single phoneme contrasts. Temporal information about voicing, although possibly redundant with changes in spectral balance, should help in the identification of consonants but not vowels. From these considerations, we would expect manipulating envelope cutoff and carrier type to have its greatest effects for sentences and least for vowels, which, in fact, appears to be the case.

The second major difference between conditions is in the density of the frequency spectrum. As shown in Fig. 1, the 3s30 condition has a sparse spectrum with broad spectral "holes" that may make it more difficult to fuse the percept into a single auditory object. One possibility is that the sidebands in the 300 Hz sine condition and the broader carrier bandwidth of the noise conditions created a denser or more continuous spectrum, which better conveys spectral shape cues. These differences in spectral density are still present after peripheral auditory filtering, as can be seen from the excitation patterns in Fig. 6 (calculated on the basis of normal auditory filtering) to three-channel vocoded versions of a neutral vowel at a fundamental frequency of 150 Hz. Such differences are still apparent for five-channel sine vocoders, although, as here, they are most prominent at lower frequencies where the auditory filters are broadest.

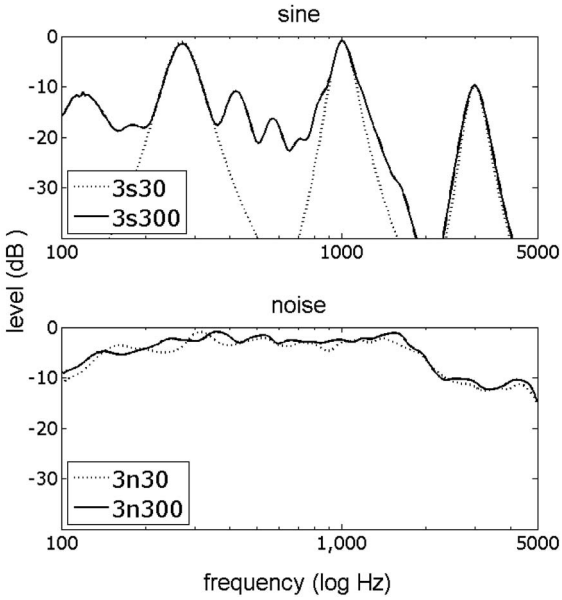


FIG. 6. Excitation patterns to three-channel vocoded versions of a synthetic static neutral vowel (formants at 500, 1500, and 2500 Hz) at a fundamental frequency of 150 Hz. These are calculated on the basis of a gammatone filter bank meant to represent normal auditory filtering (Clark, 2007). Note the lack of difference for a noise-vocoded vowel at the two envelope cutoff frequencies, but the greater degree of difference for the sine-vocoded tokens. Such differences are still apparent for five-channel sine vocoders, although, as here, they are most prominent at lower frequencies where the auditory filters are broadest. For higher numbers of channels, the differences are minor.

We might expect to see this difference in accessibility of spectral shape cues reflected in perception of consonant place, which is carried, primarily, by the frequency spectrum (Rosen, 1992). This was partially supported; place was poorest in the 30 Hz sine condition (Fig. 3) but there was also a difference between the two noise conditions and the 300 Hz sine condition, so a more continuous spectrum cannot be the only explanation.

A third difference is that the 300 Hz envelope cutoff leads to common comodulations across all the carrier bands, which may allow a more ready grouping of those components together. Several studies have shown that common modulations across the individual components of sine-wave speech can improve recognition (Carrell and Opie, 1992; Barker and Cooke, 1999; Lewis and Carrell, 2007), although this may be relatively more important when other cues to auditory object formation are not available. We address this issue in experiment 3.

III. EXPERIMENT 2: CORRELATES OF FUNDAMENTAL FREQUENCY

The purpose of experiment 2 was to determine to what extent cues to fundamental frequency (F0) were available in the various test conditions.

A. Subjects

Subjects were ten adults (eight females and two males) recruited from the student populations at UCL (seven subjects) and University of Washington, Seattle (three subjects). Six were native speakers of southern British English, one was a bilingual Swedish/English speaker, and three were native speakers of American English. None had any experience with tonal languages. Subjects ranged in age from 22 to 42 years (mean age 30 years). All had normal hearing, defined as pure-tone thresholds of 20 dB HL or better (see ANSI, 2006) at octave frequencies between 0.25 and 8 kHz. All subjects were paid for their participation.

B. Stimuli and procedure

Two sets of stimuli were used: synthetic glides and naturally produced sentences

1. Synthetic glides

The glide set included four diphthongs: /aʊ/, /eɪ/, /aɪ/, and /oɪ/. Details of stimulus creation are available in Green *et al.*, 2002. These 620-ms long tokens ranged in fundamental frequency such that the F0 at the midpoint in time of each glide was either 113 or 226 Hz. The ratio of start-to-end frequency varied in 12 equal logarithmic steps from 1:0.5 to 1:2.0, so the largest glides went from 80 to 160 Hz or from 160 to 320 Hz. The glides were vocoded using the three-band processing described for experiment 1. Final stimuli included two unprocessed conditions (F0 113 and F0 226) and eight processed conditions (two carrier types $\times$ two envelope cutoffs $\times$ two F0s).

Stimuli were presented diotically through Sennheiser HD 25 SP headphones at a comfortable listening level. On each trial the subject was required to identify the intonation as either "rise" or "fall" using a computer mouse. A block always consisted of 48 randomly ordered trials (the 4 diphthongs $\times$ 12 F0 steps). The first block in any test condition was intended as a training block and was presented with visual correct answer feedback. The remaining blocks in any test condition were presented without feedback. Subjects completed two blocks in each unprocessed condition, and four blocks in each vocoded condition. The order of test conditions was randomized, with the constraint that the subject completed the unprocessed conditions first.

2. Sentences

Sentences were drawn from 30 sentences previously used by Green *et al.* (2005). These consisted of simple declarative sentences that could be produced as either a statement or question; for example, "They're playing in the garden." The recordings were made in an anechoic room with each of the sentences read as a statement with a falling pitch contour and as a question with a rising pitch contour. One male and one female native talker of Southern British English produced the sentences. The ranges of F0 values were approximately 100–220 Hz for the male talker and 120–360 Hz for the female talker. The sentences were vocoded using the same processing as described for experiment 1. Only three-band versions were created. Final stimuli included five conditions: sine-vocoded with 30 or 300 Hz envelope cutoff, noise-vocoded with 30 or 300 Hz envelope cutoff, and unprocessed; each blocked by gender (male or female talker).

Stimuli were presented diotically through Sennheiser HD 25 SP headphones at a comfortable listening level. On a trial subjects heard a sentence and were required to identify the intonation as either "rise" or "fall" using a computer mouse. A block consisted of 10 trials (for unprocessed sentences) or 20 trials (for vocoded sentences). The first block in any test condition was intended as a training block and was presented with visual correct answer feedback. The remaining blocks in any test condition were presented without feedback. Subjects completed two blocks of the unprocessed speech and four blocks in each vocoded condition. The order of test conditions was randomized, with the restriction that each subject completed the unprocessed conditions first.

C. Results

1. Synthetic glides

Results for the glides are shown in Fig. 7. For the 3s30 and 3n30 conditions at both center F0s, and for the 3n300 condition at the 226 Hz F0, the functions are essentially flat indicating that subjects could not identify the direction of pitch change. A logistic regression was applied to the proportion of fall responses as a function of the log (base 10) of the start-to-end frequency ratio for each processing condition and center F0 for each subject in order to obtain estimates of the slopes of the functions, steeper slopes indicating better performance. Chi squared tests indicated that none of the fits deviated significantly from the observed data.[1] Slope estimates (Table II) ranged from near zero for the relatively flat functions to more than 20 for the unprocessed conditions. The slope estimates for each of the ten subjects were analyzed using a two-way ANOVA. Repeated-measures factors were processing condition and center F0. There was a significant interaction between condition and F0 ($F_{4,60}=4.80$, $p=0.007$), a significant main effect of condition ($F_{4,60}=15.70$, $p=0.001$), but no main effect of F0 ($F_{1,15}=0.35$, $p=0.565$). This is not surprising given that only in condition 3n300 are there obvious differences between the identification functions for the two frequencies. Post-hoc comparisons indicated that results were different for the 113 and 226 Hz F0s for the 3n300 condition ($t_9=2.68$, $p=0.025$) but not for the 3n30 ($t_9=-1.33$, $p=0.217$), 3s30 ($t_9=-0.25$, $p=0.806$), 3s300 ($t_9=-0.45$, $p=0.667$), or unprocessed ($t_7=0.59$, $p=0.576$) conditions.

2. Sentences

Results for the sentences are shown in Fig. 8. A two-way repeated-measures ANOVA (talker $\times$ condition) found no interaction with talker gender, $F_{4,90}=0.550$, $p=0.699$, so data are pooled across talker. There was a significant effect of condition, $F_{4,89}=39.92$, $p<0.005$. All of the conditions were different from one another ($p<0.010$) except for 3n30 and 3n300 ($p=0.081$). Like the glides, performance was best for the 3s300 condition. Unlike the glides, listeners performed above chance in almost all conditions (except 3s30),

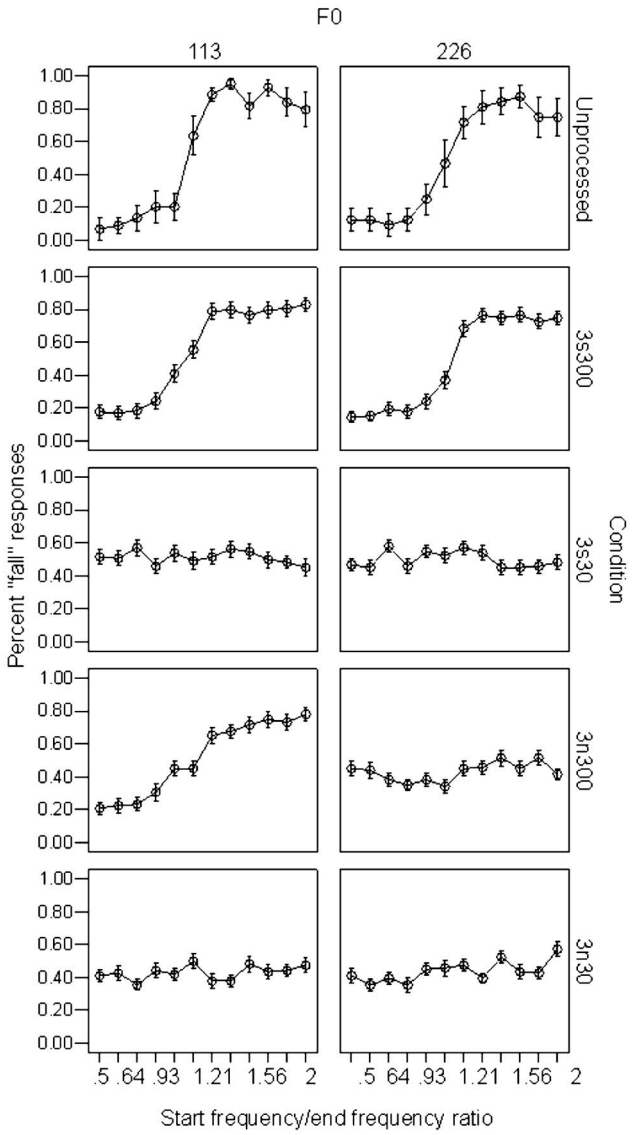


FIG. 7. Percent fall responses in various processing conditions as a function of the start-to-end frequency ratio for synthetic glides. The unprocessed condition is at top.

even for the conditions where no F0 cues to intonation had been available in the glides. Presumably, this reflects availability of other cues to intonation such as word lengthening and envelope cues to syllable (Smith, 2002). Put another way, although F0 derived from temporal envelope cues contributes to sentence intonation, other cues also play a role. In that sense, the sentence stimuli do not represent a "pure" test of the utility of F0 in judging intonation. However, they do illustrate the partial role of F0 in realistic speech materials. It is interesting that those secondary cues do not seem to be available in the 3s30 condition, where performance was close to chance. One subtle difference between question and statement that can be heard and seen in a spectrogram is the amplitude and duration of the last syllable. It is possible that the 3s30 spectrum is simply too sparse and this cue too weak to be useful.

TABLE II. Mean slope estimate (from a logistic regression) across ten subjects for the proportion of "fall" responses as a function of the start-to-end frequency ratio.

	Center F0	
Condition	113	226
3n30	0.51	0.97
3n300	9.20	0.62
3s30	−0.36	0.02
3s300	9.79	14.34
Unprocessed	22.95	20.92

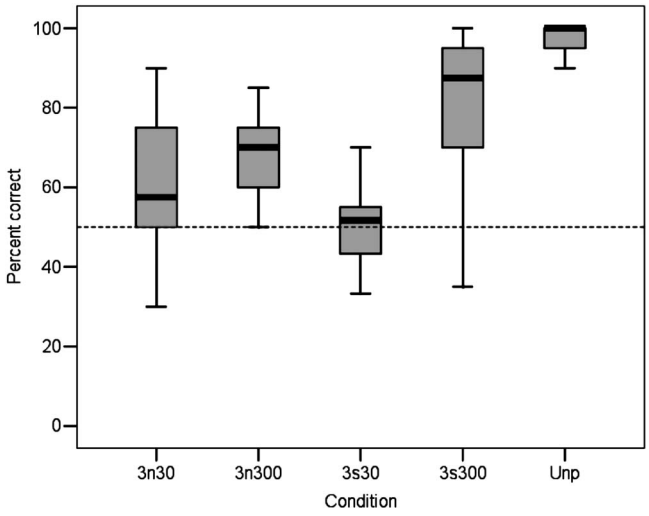


FIG. 8. Boxplot showing the range of scores, interquartile range (box length), and median for each condition in the question/statement task. Chance performance in this two-alternative forced-choice task was 50% (shown by the dashed line).

D. Discussion

Data from experiment 2 demonstrate that F0 information can be derived from envelope fluctuations as long as the envelope cutoff frequency is sufficiently high to allow for transmission of periodicity cues. The effect is largest for the sine carriers, where subjects could not derive any F0 information with a 30 Hz cutoff but performed nearly as well with the 300 Hz cutoff as they did with unprocessed stimuli.

The ability to obtain some pitch change information in the 113 Hz F0, 3n300 but not in the 226 Hz F0, 3n300 condition is consistent with previous work showing that temporal cues to voice pitch are less effective as F0 increases above about 200 Hz (Arehart and Burns, 1999; Green *et al.*, 2002, 2004; Chatterjee and Peng, 2008; Laneau *et al.*, 2006; Patterson *et al.*, 1978). This occurs because sensitivity to modulation decreases with increasing modulation frequency (Grant *et al.*, 1998).

It remains unclear to what extent F0 (versus other cues) contributed to differences in sentence, vowel, and consonant recognition seen in experiment 1. In previous work F0 made only a small contribution to sentence intelligibility (Hillenbrand, 2003) and even less for syllable-length material (Ohde, 1984; Faulkner and Rosen, 1999; Holt *et al.*, 2001). In experiment 1, we saw a large improvement in consonant and sentence intelligibility for a 300 Hz versus 30 Hz envelope cutoff. For example, there was a 40% difference in

scores between the 3s30 and 3s300 sentences. The magnitude of the improvement and the similar improvement for consonants as for sentences make it unlikely that F0 is the entire source of the difference between conditions. Accordingly, we next examine the contributions of a sparse versus continuous spectrum in combination with across-frequency comodulation.

IV. EXPERIMENT 3: CONTRIBUTIONS OF COMODULATION AND SPECTRAL DENSITY

In experiment 1, performance was best for the 300 Hz sine condition and worst for the 30 Hz sine condition. We hypothesized that this resulted because sidebands in the 300 Hz sine condition provided a denser, more continuous spectrum as opposed to the spectral "holes" prominent in the 30 Hz sine condition. Although the noise carriers would also be expected to exhibit greater spectral density compared to the 30 Hz sine condition, that (theoretical) advantage could be offset by other factors such as the random modulations of the noise carrier itself, or to spectral smearing due to overlap between the noise bands.

Another difference between conditions was the degree to which information was comodulated across bands. Common amplitude modulation may "cohere" the output of separate acoustic filters into a single auditory object. We hypothesized that listeners would have more difficulty cohering the sparse information in the 30 Hz sine condition and this might have contributed to performance differences. Because comodulation and spectral shape definition covaried in the original conditions, the purpose of experiment 3 was to create test conditions that varied the degree of comodulation separately from spectral density.

A. Subjects

Subjects were 15 adults (8 females and 7 males) recruited from the student populations at UCL (11 subjects, native speakers of southern British English) and University of Washington, Seattle (4 subjects, native speakers of American English). One subject had participated in experiment 1 and the remaining subjects had no prior experience with the test materials. Subjects ranged in age from 19 to 62 years (mean 30 years). All but two listeners had normal hearing, defined as pure-tone thresholds of 20 dB HL or better (see ANSI, 2006) at octave frequencies between 0.25 and 8 kHz. The two oldest subjects aged 54 and 62 years met the criteria through 4 kHz but had mild (40 dB HL or better) loss at 8 kHz. All subjects were paid for their participation.

B. Stimuli and procedure

Consonant recognition was measured with multiple tokens of the 20 vowel-consonant-vowel utterances used in experiment 1, spoken by a male and a female talker who were native speakers of southern British English. The female talker was the same speaker as for experiment 1. Syllables were produced with the consonants in three different vocalic contexts (/ɑ/, /iː/, and /uː/ hence "ah," "ee," and "oo"). From these, we selected six tokens of each combination of vowel and consonant, including three from the male talker and three from the female. Across all exemplars, the mean F0 for the male talker was 91 Hz and the mean F0 for the female talker was 203 Hz.

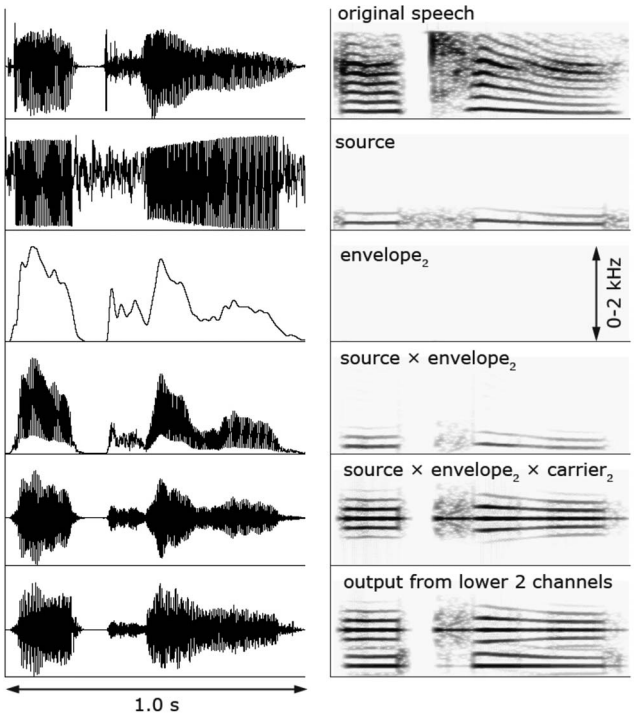


FIG. 9. Time waveforms and narrow-band spectrograms illustrating the various stages in the construction of stimuli in the Px condition of experiment 3. No spectrogram is shown for the envelope signal as it only contains a narrow band of low frequencies. At top is shown the original VCV /ɑkɑ/ as spoken by the female talker. Row 2 shows the source function that has a periodic wave matching that of the speech when it is voiced, and a random noise otherwise. Row 3 shows the envelope extracted from channel 2 (hence the subscript), the middle channel of this three-channel vocoding. Row 4 shows the result of multiplying the given envelope by the source, and row 5 the result of a further multiplication by the sine carrier appropriate for this channel. Finally, row 6 shows the final output of channels 1 and 2 summed. The third (highest) channel is not visible because the frequency range of the spectrograms has been limited to 2 kHz so that the harmonics of the voiced source can be clearly observed.

Four test conditions used three-band vocoding with either a sine or noise carrier, and either a 300 or 30 Hz envelope cutoff. These were essentially identical to experiment 1 except that we used multiple utterances of the same token, and a male talker as well.

Three further conditions were created to vary comodulation separately from spectral density, all based on three-channel vocoders and a synthetic source. Figure 9 exhibits the various stages in the process. To create a source signal for voiced speech, laryngograph recordings were used to determine the time points of individual vocal fold closures, referred to here as pitch pulses. A speech analysis program (SFS)[2] was used to generate a fundamental-frequency contour based on the pitch pulses, which was used along with the original waveform for hand correction of any pitch pulse errors. We then created a sawtooth wave whose periods corresponded to those of the pitch pulses (i.e., which varied in F0 as did the original signal). For each band, the fundamental-frequency modulated sawtooth carrier was filtered to constrain its spectral slope and bandwidth to the

same 300 Hz span that would have occurred for a typically obtained envelope signal using this cutoff. The source for all other intervals of speech was a filtered random noise (i.e., during voiceless speech or silence). See row 2 of Fig. 9 for an example of the source function on its own.

Each speech file was then digitally filtered into three bands, using the same filter bank as previously described. The output of each band was full-wave rectified and low-pass filtered at 30 Hz (fourth-order Butterworth) to extract the amplitude envelope (Fig. 9, row 3 shows the envelope for the middle channel). The source wave was then multiplied by the 30 Hz envelope for each band (Fig. 9, row 4), and the result used to modulate a sine carrier of the appropriate frequency (Fig. 9, row 5). The rms level was adjusted at the output of the filter to match the original analysis and the signal was summed across bands (Fig. 9, row 6). This signal was a control condition termed Px (pulsed excitation), and should be very similar to the 3s300 condition, as can be seen in Fig. 10. One Px condition was created for the male talker and one for the female talker.

A "decoherent" (Dx) signal was created using a constant pitch contour, which varied across channels. For the male speech, the rate was set to 90.7, 79.7, and 110.9 Hz in bands 1–3, respectively, and for the female speech, 202.7, 178.3, and 247.7 Hz in bands 1–3, respectively. The different sawtooth rates were modeled on work by Carrell and Opie (1992) and, as in that study, values were chosen to prevent short-period unintentional comodulation due to a large common factor across the three frequencies.

In order to rule out the possibility that any decrements in performance for the Dx signal resulted from the flattening of the pitch contour as opposed to the differential pulse rates across channels, two sets of monotone signals (Mx) were created using a fixed pulse rate in each channel (91 Hz for the male talker and 203 Hz for the female talker). An example of a stimulus from both the Dx and Mx conditions can be found in Fig. 10.

Test procedures were the same as described for experiment 1 with the following modifications. (1) The practice block consisted of 40 trials, in which the 20 consonants were sampled from the various vowel contexts and test conditions. (2) In the test phase, each block contained 60 trials (20 consonants × 3 vowel contexts). A single block contained stimuli for either the male or female talker. (3) A trial was randomly drawn from the three available exemplars for that talker/consonant/vowel token. (4) The order of test conditions was randomly selected for each subject.

C. Results and discussion

Results, shown in Fig. 11, were analyzed with a two-way (talker × condition) repeated-measures ANOVA. There was no significant effect of talker ($F_{1,28}=2.50$, $p=0.13$), no interaction between talker and test condition ($F_{6,168}=0.78$, $p=0.52$), but a significant effect of test condition ($F_{6,168}=21.79$, $p<0.01$). Post-hoc means comparisons indicated that there was no difference between the Dx and Mx ($t_{29}=0.96$, $p=0.35$), Px and Mx ($t_{29}=-1.40$, $p=0.17$), or Dx and Px ($t_{29}=-0.56$, $p=0.58$) conditions. That is, decohering the spectrum had no significant effect when the density of the spectrum was maintained. This may be thought surprising, since the value of a common (but not independent) modulation of sinusoidal components has been demonstrated previously (Carrell and Opie, 1992). However, in a follow-up study, Lewis and Carrell (2007) also found just as much benefit for independent modulations of the sinusoidal components as for common ones. They noted that the inclusion of sounds such as medial voiceless plosives would cause comodulations across bands (although at lower modulation frequencies), which might have conferred their own cues for auditory grouping, thereby reducing the need for (and benefit of) comodulation due to voice pitch variations. Certainly all of those elements would have been present in our naturally produced speech.

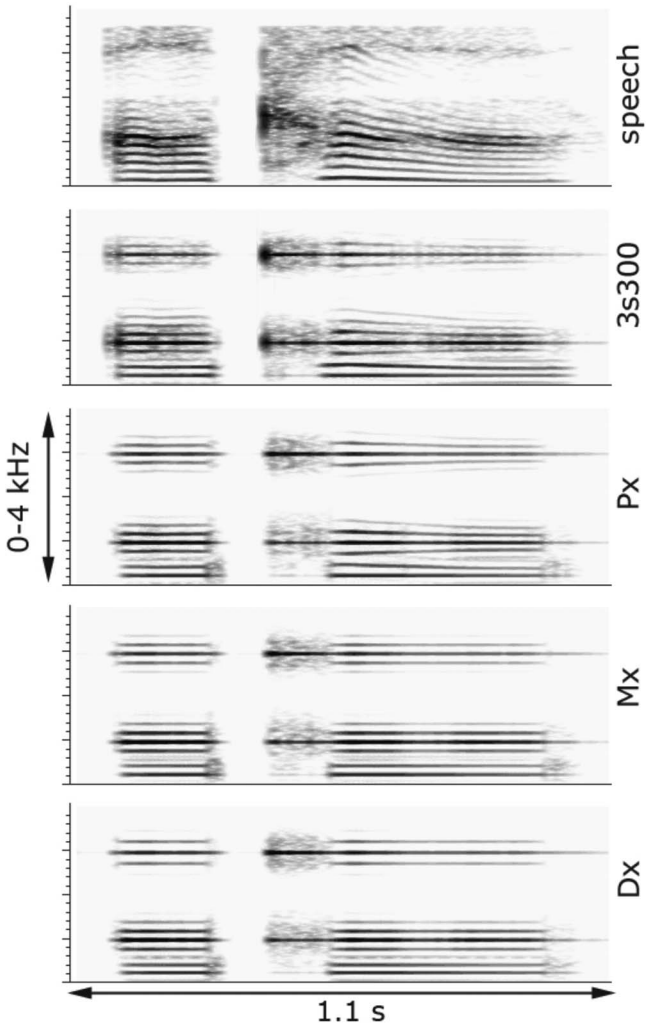


FIG. 10. Narrow-band spectrograms of versions of the VCV /ɑkɑ/ as spoken by the female talker of experiment 3. The top panel shows the original speech token, followed by the conditions of particular interest (for details, see the text). Note the similarity between conditions 3s300 and Px. Condition Mx is similar to Px, but with a fixed F0, identical for all three channels, as can be seen from the fact that all the spectral components are fixed in frequency (unlike the varying ones in Px). Condition Dx also uses fixed F0s, but ones which vary from channel to channel (note that the spectral components in the highest frequency channel are more widely spaced than those in the lower two). Also of interest is the fact that the transient release burst of the /k/, shown prominently by the dark "blobs" for both the speech and 3s300 sounds at consonantal onset, is more or less eliminated by the 30 Hz envelope filter used in the other conditions.

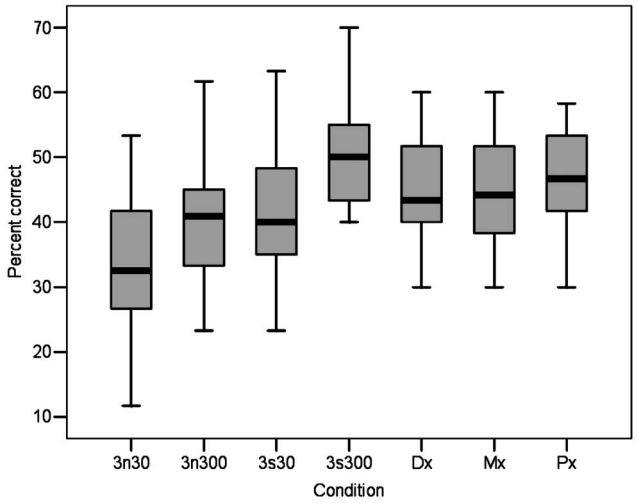


FIG. 11. Boxplot showing the range of scores, interquartile range (box length), and median for each condition in a consonantal identification task.

Although the Px processing was intended to mimic the properties of the 300 Hz sine condition, scores were about 5% higher in 3s300 than in Px ($t_{29}=3.57$, $p<0.01$). The reasons for this are apparent in Table III, which shows confusion matrices for the two conditions collapsed across manner/voicing categories. Performance for voiced fricatives is much better for 3s300, almost certainly attributable to the fact that the Px processing does not allow the representation of mixed excitation (having both a periodic and a noise source) as happens in voiced fricatives such as /ʒ/. Plosives are also better identified in 3s300, probably because the 30 Hz smoothing filter in the Px processing makes the onset burst of the consonant less prominent, as can be seen in the spectrograms in Fig. 9. The only sounds for which Px scores are substantially higher than 3s300 are the nasals but this is balanced out almost exactly by the better performance for glides by 3s300. It is not clear why this difference arises between the two conditions, since the representation of periodicity for such sounds is unlikely to differ much. But glides and nasals are the two classes most often confused in both processing conditions, so may best be thought of as a single category.

TABLE IV. Mean percent correct for experiments 1 and 3, and for only the /ɑ/ vowel context in experiment 3. 16 subjects participated in experiment 1 and 15 (different) subjects participated in experiment 3.

Condition	Experiment 1	Experiment 3 /ɑ/ only	Experiment 3 all vowels
Noise, 30 Hz	52.4	46.0	35.3
Noise, 300 Hz	50.8	54.0	43.3
Sine, 30 Hz	48.3	44.3	41.4
Sine, 300 Hz	58.4	50.7	53.1

As in experiment 1, scores were highest for the 300 Hz sine condition, with lower scores for the 300 and 30 Hz noise conditions and the 30 Hz sine condition. That scores are lower when all vowels were included is not surprising, because more token variability typically leads to lower scores. Table IV compares mean scores for experiment 1, for all vowels in experiment 3, and for only the /ɑ/ vowel context in experiment 3. When only the /ɑ/ vowel is considered, mean scores are generally similar (within 5%) across experiments. The largest difference is the mean score for the 300 Hz sine condition, which is 8% higher in experiment 1. The reasons for this are unclear but in experiment 1 subjects might also

TABLE III. Comparison of confusion matrices collapsed across manner and voicing for the Px (top) and 300 Hz sine (bottom) conditions in experiment 3. For ease of reading, any cell with less than five responses is rendered as blank. "v+" indicates sounds that are voiced and "v−" sounds that are voiceless.

		Response								
Px		v+ plosive	v− affricate	v− fricative	v+ affricate	v− plosive	Glide	Nasal	v+ fricative	Sum
Stimulus	v+ plosive	235			9	17				270
	v− affricate		83			5				90
	v− fricative			255					5	270
	v+ affricate	7			81					90
	v− plosive		40			224				270
	Glide	14					204	60	80	360
	Nasal						46	122	8	180
	v+ fricative	12			10		57		183	270
		Response								
3s300		v+ plosive	v− affricate	v− fricative	v+ affricate	v− plosive	Glide	Nasal	v+ fricative	Sum
Stimulus	v+ plosive	246				13				270
	v− affricate		77		8					90
	v− fricative			253					5	270
	v+ affricate	10			79					90
	v− plosive		26			239				270
	Glide	5					220	58	73	360
	Nasal						66	104	8	180
	v+ fricative	13			20		24		209	270

TABLE V. Mean percent correct for experiments 1 and 3, and for only the /ɑ/ vowel context in experiment 3. Results are for the same five subjects for both experiments.

Condition	Experiment 1	Experiment 3 /ɑ/ only	Experiment 3 all vowels
Noise, 30 Hz	55.6	46.0	31.1
Noise, 300 Hz	51.2	56.0	38.8
Sine, 30 Hz	47.2	44.0	41.8
Sine, 300 Hz	62.9	57.0	48.5

have heard the processed vowels and/or processed sentences before the consonants. It is possible that familiarization affected the more intelligible stimuli to a greater extent. Whitmal *et al.* (2007) also noted the likelihood of such "exposure" effects for their vocoded stimuli.

Other findings replicated the results of experiment 1: For the tone carrier, scores were higher for a 300 Hz than a 30 Hz envelope cutoff ($t_{29}=-5.85$, $p<0.01$); for the 300 Hz cutoff, scores were higher for a tone than for a noise ($t_{29}=-7.13$, $p<0.01$). Those results were verified using a new set of five subjects with normal hearing who completed both experiments, and for whom order of the two experiments was counterbalanced across subjects. Data are shown in Table V and again, show the lowest score for the sine 30 Hz condition when only /ɑ/ vowels were used, and the lowest score for the noise 30 Hz condition when all vowels were included.

V. GENERAL DISCUSSION

Although the differences were most pronounced for the sentences, all materials shared common patterns. With a 30 Hz envelope cutoff frequency, performance was better with a noise than with a sine carrier. With a 300 Hz envelope cutoff frequency, performance was better with a sine than a noise carrier. For sine-vocoded speech, performance was better with a 300 Hz than a 30 Hz envelope cutoff. For noise-vocoded speech, there was no benefit of increasing the envelope cutoff.

To what should we attribute these effects? First, consider the effect of increasing envelope cutoff frequency. According to Rosen's (1992) classification, a 30 Hz envelope cutoff would not transmit cues concerning periodicity (or aperiodicity). Therefore, detection of voicing should be better with the 300 Hz cutoff, and this is supported by the feature analysis.

The higher-frequency envelope modulations would also be expected to provide cues to variations in F0, which can contribute to sentence recognition in quiet and to a smaller extent to consonant recognition in quiet (Faulkner *et al.*, 2000). For example, F0 can code some segmental characteristics such as plosive consonant voicing and aspiration (Haggard *et al.*, 1970). In tonal languages such as Mandarin Chinese, F0 variations are more important than in English (Xu *et al.*, 2002; Xu and Pfingst, 2003). F0 cues will also be important in background noise where they can serve as a cue to talker separation. Stone *et al.* (2008) demonstrated that increasing the envelope cutoff frequency from 45 to 180 Hz improved sentence recognition in a competing speech task even when the signal already offered a high level of spectral detail via a large number of bands. In contrast to the present data, Stone *et al.* (2008) also found some (although small) benefit to increasing envelope filter cutoff for noise-vocoded as well as sine-vocoded signals. This may be related to the nature of the speech-in-noise task as opposed to recognition of speech in quiet; that is, the higher-frequency envelope modulations might have aided talker separation even if not improving recognition *per se*.

Is the advantage of increasing the envelope cutoff frequency with a sine carrier related to the information conveyed by the spectral sidebands, or to the availability of higher-frequency temporal modulations? Gonzalez and Oliver (2005) believed that the ability to identify talker gender using fewer bands with sine-wave carriers than with noise band carriers was partially due to the ability to resolve spectral sidebands for the sine carrier (reflecting the spectral content of the envelope), which of course would signal talker F0, a strong cue to gender. If the effect was solely due to a denser (or more continuous) spectrum, though, we would expect to see similar performance for the 300 Hz sine and 300 Hz noise conditions, yet the sine was much superior. Nor can we attribute the better performance with the 300 Hz sine condition to a greater sense of comodulation across frequency bands, since "decohering" a signal to abolish across-band asynchrony had no effect on recognition. It therefore appears that the transmission of better cues to periodicity/aperiodicity (signaling voicing) and its variations in F0 (signaling intonation) are the main contributors to this advantage.

Why didn't increasing the envelope cutoff frequency confer the same benefit for the noise carrier? Unlike a constant-amplitude sine carrier, a noise carrier has its own envelope fluctuations. These fluctuations could act as interferers, preventing full use of envelope information (Gonzalez and Oliver, 2005), and perhaps negating the benefit of the higher-rate amplitude modulations. Such an effect has already been demonstrated for nonspeech signals (Dau *et al.*, 1999). The idea is that listeners will have more trouble detecting a signal in a randomly-varying masker because the auditory system will not be able to easily distinguish between a signal, and a randomly occurring component of the masker. For whatever reasons, it is well known that both the detection of amplitude modulations in noise carriers and discrimination of modulation rates worsen with increasing modulation rates (Burns and Viemeister, 1976, 1981; Patterson *et al.*, 1978; Viemeister, 1979). A "low-noise noise" can be created by restricting the relationship between noise components such that they will have related phase. Whitmal *et al.* (2007) demonstrated better performance when the signal was vocoded using either sine carriers or low-noise noise carriers, and worse performance with conventional noise carriers. The worst performance was with a narrow-band Gaussian noise carrier, which would have had the greatest inherent (and potentially interfering) envelope fluctuations. A similar

concept might explain the restriction of the benefit of increasing envelope cutoff frequency from 30 to 300 Hz with a noise carrier.

It is also possible that higher-rate temporal fluctuations (i.e., as occur with the higher envelope cutoff frequency) are valuable only when compared to the case where spectral cues alone are insufficient. Xu and co-workers (2005, 2007, 2008) noted a tradeoff such that recognition can be improved either by increasing the envelope cutoff frequency or by increasing the number of spectral bands. In this instance, the 30-Hz noise condition offered about the same level of spectral detail as the 300-Hz noise condition. Perhaps listeners rely more heavily on spectrum to aid identification and the additive effect of higher temporal fluctuations is simply too subtle to contribute in a measurable way.

In summary, our results indicate that changes in carrier type in combination with envelope filter cutoff can significantly alter available cues in vocoded speech. Spectral density and the availability of higher-rate temporal modulations were indicated as major factors. Subjects performed worst for conditions that were spectrally sparse and contained only low-rate temporal modulations. Based on the results of experiment 3, comodulation across bands appeared to make a negligible contribution, at least for these speech materials.

These results also suggest that one should give careful thought to the choice of vocoding method for applications such as cochlear implant simulations. First, consider the level of spectral information provided. In this case, although the number of bands was chosen in part to prevent floor and ceiling effects for our speech tasks, it is within the range of the four to eight effective channels expected in real CI users (Wilson and Dorman, 2008). However, high envelope cut-off frequencies combined with sinusoidal carriers lead to sensitivity to voice pitch variations in normal listeners that far outstrips the performance of any traditional cochlear implant user.

It may also be the case that different simulations mimic different aspects of cochlear implants, but that there is no single implementation that can simulate all aspects and situations. Tone vocoders with high cutoff frequencies are not representative of real implants, but a low envelope cutoff probably results in worse sensitivity to voice pitch than in an actual cochlear implant. Noise-vocoding with high cutoffs probably has appropriate sensitivity to voice pitch, but the fluctuations from the noise itself would not be a factor in a real implant. Although it may not be possible to create a situation where a normal-hearing listener responds as would a cochlear implant wearer in all conceivable ways, different implementations of vocoding can lead to more-or-less realistic results.

ACKNOWLEDGMENTS

The authors thank Tim Green for providing the stimuli for experiment 2, and for many helpful conversations about experimental design and analysis. Many thanks to Sam Eaton-Rosen for crucial help in creating Figs. 1, 9, and 10. The support of NIDCD (Grant No. R01 DC006014) and the Bloedel Hearing Research Foundation to P.S. are gratefully acknowledged, as is the support of the UK Royal National Institute for Deaf People (RNID) to S.R.

[1]Degrees of freedom were either 22 or 46 depending on the number of trials in the block.

[2]Speech Filing System SFS-WIN v 1.5, http://www.phon.ucl.ac.uk/resource/sfs/ (Last viewed 5/14/2009).

ANSI (**2006**). American National Standards Institute S3.22-2006, New York.

Arehart, K. H., and Burns, E. M. (**1999**). "A comparison of monotic and dichotic complex-tone pitch perception in listeners with hearing loss," J. Acoust. Soc. Am. **106**, 993–997.

Barker, J., and Cooke, M. (**1999**). "Is the sine-wave speech cocktail party worth attending?," Speech Commun. **27**, 159–174.

Bench, J., and Bamford, J. (**1979**). *Speech Hearing Tests and the Spoken Language of Hearing Impaired Children* (Academic, London).

Bor, S., Souza, P., and Wright, R. (**2008**). "Multichannel compression: Effects of reduced spectral contrast on vowel identification," J. Speech Lang. Hear. Res. **51**, 1315–1327.

Burns, E. M., and Viemeister, N. F. (**1976**). "Nonspectral pitch," J. Acoust. Soc. Am. **60**, 863–869.

Burns, E. M., and Viemeister, N. F. (**1981**). "Played-again SAM: Further observations on the pitch of amplitude-modulated noise," J. Acoust. Soc. Am. **70**, 1655–1660.

Carrell, T. D., and Opie, J. M. (**1992**). "The effect of amplitude comodulation on auditory object formation in sentence perception," Percept. Psychophys. **52**, 437–445.

Chang, Y., and Fu, Q.-J. (**2006**). "Effects of talker variability on vowel recognition in cochlear implants," J. Speech Lang. Hear. Res. **49**, 1331–1341.

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

Clark, N. (**2007**). GAMMATONE TOOL KIT v2.0 (MathWorks), http://www.mathworks.com/matlabcentral/fileexchange/15313 (Last viewed 5/14/2009).

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.

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.

Faulkner, A., and Rosen, S. (**1999**). "Contributions of temporal encodings of voicing, voicelessness, fundamental frequency, and amplitude variation to audio-visual and auditory speech perception," J. Acoust. Soc. Am. **106**, 2063–2073.

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

Fu, Q. J., and Shannon, R. V. (**2000**). "Effect of stimulation rate on phoneme recognition by nucleus-22 cochlear implant listeners," J. Acoust. Soc. Am. **107**, 589–597.

Gallun, F., and Souza, P. (**2008**). "Exploring the role of the modulation spectrum in phoneme recognition," Ear Hear. **29**, 800–813.

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.

Grant, K. W., Summers, V., and Leek, M. R. (**1998**). "Modulation rate detection and discrimination by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **104**, 1051–1060.

Green, T., Faulkner, A., and Rosen, S. (**2002**). "Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants," J. Acoust. Soc. Am. **112**, 2155–2164.

Green, T., Faulkner, A., and Rosen, S. (**2004**). "Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants," J. Acoust. Soc. Am. **116**, 2298–2310.

Green, T., Faulkner, A., Rosen, S., and Macherey, O. (**2005**). "Enhancement of temporal periodicity cues in cochlear implants: Effects on prosodic perception and vowel identification," J. Acoust. Soc. Am. **118**, 375–385.

Green, T., Katiri, S., Faulkner, A., and Rosen, S. (**2007**). "Talker intelligibility differences in cochlear implant listeners," J. Acoust. Soc. Am. **121**,

EL223–229.
Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.
Haggard, M., Ambler, S., and Callow, M. (**1970**). "Pitch as a voicing cue," J. Acoust. Soc. Am. **47**, 613–617.
Hillenbrand, J. (**2003**). "Some effects of intonation contour on sentence intelligibility," J. Acoust. Soc. Am. **114**, 2338.
Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2001**). "Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?," J. Acoust. Soc. Am. **109**, 764–774.
Laneau, J., Moonen, M., and Wouters, J. (**2006**). "Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants," J. Acoust. Soc. Am. **119**, 491–506.
Lewis, D. E., and Carrell, T. D. (**2007**). "The effect of amplitude modulation on intelligibility of time-varying sinusoidal speech in children and adults," Percept. Psychophys. **69**, 1140–1151.
Loebach, J. L., and Wickesberg, R. E. (**2006**). "The representation of noise vocoded speech in the auditory nerve of the chinchilla: Physiological correlates of the perception of spectrally reduced speech," Hear. Res. **213**, 130–144.
MacLeod, A., and Summerfield, Q. (**1990**). "A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use," Br. J. Audiol. **24**, 29–43.
Mauchly, J. W. (**1940**). "Significance test for sphericity of a normal n-variate distribution," Ann. Math. Stat. **11**, 204–209.
Ohde, R. N. (**1984**). "Fundamental frequency as an acoustic correlate of stop consonant voicing," J. Acoust. Soc. Am. **75**, 224–230.
Patterson, R. D., Johnson-Davies, D., and Milroy, R. (**1978**). "Amplitude-modulated noise: The detection of modulation versus the detection of modulation rate," J. Acoust. Soc. Am. **63**, 1904–1911.
Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London **336**, 367–373.
Rosen, S., Faulkner, A., and Wilkinson, L. (**1999**). "Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants," J. Acoust. Soc. Am. **106**, 3629–3636.
Shannon, R. V., Fu, Q. J., and Galvin, J., III (**2004**). "The number of spectral channels required for speech recognition depends on the difficulty of the listening situation," Acta Oto-Laryngol. **552**, 50–54.
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Smith, C. L. (**2002**). "Prosodic finality and sentence type in French," Lang Speech **45**, 141–178.
Souza, P., and Boike, K. T. (**2006**). "Combining temporal-envelope cues across channels: Effects of age and hearing loss," J. Speech Lang. Hear. Res. **49**, 138–149.
Souza, P., and Turner, C. W. (**1998**). "Multichannel compression, temporal cues and audibility," J. Speech Lang. Hear. Res. **41**, 315–326.
Stone, M. A., Fullgrabe, C., and Moore, B. C. J. (**2008**). "Benefit of high-rate envelope cues in vocoder processing: Effect of number of channels and spectral region," J. Acoust. Soc. Am. **124**, 2272–2282.
Turner, C. W., Souza, P. E., and Forget, L. N. (**1995**). "Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners," J. Acoust. Soc. Am. **97**, 2568–2576.
Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.
Whitmal, N. A., Poissant, S. F., Freyman, R. L., and Helfer, K. S. (**2007**). "Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience," J. Acoust. Soc. Am. **122**, 2376–2388.
Wilson, B. S., and Dorman, M. F. (**2008**). "Cochlear implants: A remarkable past and a brilliant future," Hear. Res. **242**, 3–21.
Xu, L., and Pfingst, B. E. (**2003**). "Relative importance of temporal envelope and fine structure in lexical-tone perception (L)," J. Acoust. Soc. Am. **114**, 3024–3027.
Xu, L., and Pfingst, B. E. (**2008**). "Spectral and temporal cues for speech recognition: Implications for auditory prostheses," Hear. Res. **242**, 132–140.
Xu, L., Thompson, C. S., and Pfingst, B. E. (**2005**). "Relative contributions of spectral and temporal cues for phoneme recognition," J. Acoust. Soc. Am. **117**, 3255–3267.
Xu, L., Tsai, Y., and Pfingst, B. E. (**2002**). "Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses," J. Acoust. Soc. Am. **112**, 247–258.
Xu, L., and Zheng, Y. (**2007**). "Spectral and temporal cues for phoneme recognition in noise," J. Acoust. Soc. Am. **122**, 1758–1764.

Sensitivity to interaural time difference with bilateral cochlear implants: Development over time and effect of interaural electrode spacing

Becky B. Poon
Cochlear Implant Research Laboratory, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114 and Speech and Hearing Bioscience and Technology Program, Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Donald K. Eddington
Cochlear Implant Research Laboratory, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114; Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts 02114; and Speech and Hearing Bioscience and Technology Program, Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Victor Noel
Cochlear Implant Research Laboratory, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114

H. Steven Colburn
Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215

(Received 22 May 2008; revised 6 May 2009; accepted 15 May 2009)

Sensitivity to interaural time difference (ITD) in constant-amplitude pulse trains was measured in four sequentially implanted bilateral cochlear implant (CI) subjects. The sensitivity measurements were made as a function of time beginning directly after the second ear was implanted, continued for periods of months before subjects began wearing bilateral sound processors, and extended for months while the subjects used bilateral sound processors in day-to-day listening. Measurements were also made as a function of the relative position of the left/right electrodes. The two subjects with the shortest duration of binaural deprivation before implantation demonstrated ITD sensitivity soon after second-ear implantation (before receiving the second sound processor), while the other two did not demonstrate sensitivity until after months of daily experience using bilateral processors. The interaural mismatch in electrode position required to decrease ITD sensitivity by a factor of 2 (half-width) for CI subjects was five times greater than the half-width for interaural carrier-frequency disparity in normal-hearing subjects listening to sinusoidally amplitude-modulated high-frequency tones. This large half-width is likely to contribute to poor binaural performance in CI users, especially in environments with multiple broadband sound sources.
 [DOI: 10.1121/1.3158821]

PACS number(s): 43.66.Ts, 43.66.Pn [JCM] Pages: 806–815

I. INTRODUCTION

The advantages of binaural over monaural listening for normal-hearing subjects include more accurate sound-source localization (Oldfield and Parker, 1986) and superior reception of a target sound when spatially separated from competing sound sources (Zurek, 1993). It is not, therefore, surprising that hearing-impaired patients have undergone bilateral implantation in an effort to restore some measure of these binaural advantages.

Multiple investigators have tested the ability of cochlear implant (CI) users to localize sound sources using their commercial sound-processing systems (e.g., van Hoesel and Tyler, 2003; Nopp *et al.*, 2004; Litovsky *et al.*, 2006b; Grantham *et al.*, 2007) and to receive speech in the presence of spatially separated interfering sounds (e.g., Schleich *et al.*, 2004; Litovsky *et al.*, 2006a; Ricketts *et al.*, 2006; Peters *et al.*, 2007). While these studies show better performance for bilateral than monolateral listening, the level of bilateral advantage does not typically reach that of normal-hearing listeners. Furthermore, there is evidence that almost all of the benefits for binaural CI users are due to interaural level differences (ILDs), while little or no functional benefit has been shown from use of interaural time difference (ITD).

For example, two studies measured the total root-mean-squared error for localizing sound sources in the frontal horizontal plane for both CI users (listening with commercial sound-processing systems) and normal-hearing subjects using the same methods (Poon, 2006; Grantham *et al.*, 2007). The mean error scores for the CI users were 28° and 30°; substantially larger (worse) than the 7° and 0° measured for the normal-hearing listeners. Several lines of evidence suggest that CI patients use ILD as the main cue for localization. van Hoesel (2004) pointed out the correspondence between localization error magnitude and ILD cue ambiguity as a function of azimuth. Grantham *et al.* (2007) demonstrated that the impact of a sound's spectral content on CI users'

ability to localize was consistent with ILD being the principal cue. Poon (2006) and Grantham *et al.* (2008) used a single acoustic signal to measure both localization error and sensitivity to ITD and ILD in the same subjects and analyzed the degree to which the variation in ITD and ILD sensitivity could account for the variation in localization error. Both concluded that localization performance was based primarily on ILD cues. Seeber and Fastl (2008) studied two subjects with relatively good localization ability to assess the impact of several signal manipulations (spectral/temporal composition, presence/absence of the head-shadow cue, and virtual shifting of ITD/ILD cues) on the subjects' abilities to localize with their commercial sound processors; they concluded that both subjects relied predominantly on ILD cues, with little evidence of any contribution from ITD cues.

Poor ITD sensitivity is also implicated as a factor limiting the binaural advantage associated with speech reception in the presence of spatially separated speech-shaped noise sources. Zurek's (1993) model analysis of this advantage measured in normal-hearing listeners concluded that it can largely be accounted for by two components: improvement in the signal-to-noise ratio at one ear due to the purely acoustic head-shadow effect and the binaural unmasking predicted solely from ITD. A third, relatively small, component of binaural advantage is the effect of listening to the better signal-to-noise ratio signal with both ears rather than just the better ear alone. This is called the "diotic effect" because the same signal is delivered to both ears. When measured in terms of speech-reception threshold in normal-hearing subjects and when the target is straight ahead and the masker is on the sides, the head-shadow effect provides an advantage of about 8 dB, the binaural unmasking ranges from approximately 2.5 to 5 dB, and the diotic effect ranges from 1.5 to 2.5 dB (Bronkhorst and Plomp, 1988; Arsenault and Punch, 1999). In the case of adult CI users listening through their sound processors, the advantages are smaller. Measurements of the head-shadow advantage range from 5.6 to 6.8 dB, and the "squelch" advantage, which combines the benefit from ITD and diotic listening, ranges from approximately 0.9 to 1.9 dB (Schleich *et al.*, 2004; Litovsky *et al.*, 2006a). van Hoesel *et al.* (2008) measured diotic benefit for CI users to be about 1.0–1.5 dB but found no binaural unmasking. The absence of binaural unmasking and the close agreement between the van Hoesel *et al.* (2008) measures of diotic benefit and measures of squelch benefit (Schleich *et al.*, 2004; Litovsky *et al.*, 2006a) are consistent with CI users not being able to access the information normally carried by ITD. The results of the study of van Hoesel *et al.* (2008) found similar results using both commercial sound-processing strategies that discard fine-timing information and a research processor that explicitly coded fine-timing cues.

Psychophysical measures of implanted subjects' sensitivities to ITD and ILD in conditions that eliminate the possible adverse impact of commercial sound-processing strategies and multichannel stimulation are also consistent with the hypothesis that poor ITD sensitivity largely accounts for the abnormally small localization and binaural squelch advantages seen in CI users. Just noticeable differences (JNDs) for electric ILD measured using unmodulated pulse trains stimulating single interaural electrode pairs are as small as 0.2 dB (e.g., van Hoesel and Tyler, 2003). Acoustic (through-processor) ILD thresholds are near normal, with JNDs of 1–2 dB SPL (sound pressure level) measured for the best users (Laback *et al.*, 2004). In contrast, ITD JNDs measured in CI subjects are substantially worse than normal. ITD sensitivity is best for low-rate (40–100 pps) pulse trains where ITD JNDs of 50 μs or less have been measured for some interaural electrode pairs in the best subjects but typically range between 100 and 500 μs, with JNDs for a few subjects extending beyond 700 μs (van Hoesel *et al.*, 1993; van Hoesel and Clark, 1997; Lawson *et al.*, 1998; van Hoesel *et al.*, 2002; van Hoesel and Tyler, 2003; Wilson *et al.*, 2003; van Hoesel, 2004; Laback *et al.*, 2007; van Hoesel, 2007). As the pulse repetition rate increases, ITD sensitivity rapidly decreases, with JNDs typically increasing to 400 μs or greater by 600 pps (van Hoesel and Tyler, 2003; Majdak *et al.*, 2006; van Hoesel, 2007). This effect of rate on ITD is similar to that for normal-hearing subjects listening to high-frequency, bandpass-filtered acoustic clicks (Hafter and Dye, 1983), but unlike normal-hearing subjects listening to pure tones, when the ITD JND decreases from approximately 75 μs at 90 Hz to 11 μs at 1000 Hz (Klumpp and Eady, 1956).

In this paper, results are presented that further characterize the ITD sensitivity of single-interaural electrode pairs stimulated with unmodulated pulse trains in bilaterally implanted CI subjects. Though only four CI listeners participated in the current study, their results provide data over extended periods of time and are relevant for several questions that have not been fully addressed and that are of current interest. The first set of results addresses whether the factors of (adult-onset) binaural deprivation and listening experience with bilateral CIs impact ITD sensitivity in adult CI users. Several studies suggest that deprivation and listening experience influence adult monolateral (Tyler and Summerfield, 1996; Pelizzone *et al.*, 1999) and childhood bilateral (Peters *et al.*, 2007) performance as well as the development of both cortical (Sharma *et al.*, 2005) and brainstem (Gordon *et al.*, 2007) potentials in young children. Except for a report showing that localization performance did not improve for most adult subjects between 5 and 15 months of post-activation (Grantham *et al.*, 2007), the impact of deprivation and listening experience on binaural advantages or on ITD and ILD sensitivity has not been studied. While the results presented in this paper were not generated in a study explicitly designed to directly address the impact of deprivation and listening experience on ITD sensitivity in postlingually deafened CI users and are not, therefore, optimal in terms of some study parameters (see Sec. II), they are of interest as the first data of their kind.

While the potential desirability of stimulating interaurally place-matched electrodes was recognized at the beginning of bilateral cochlear implantation (e.g., van Hoesel *et al.*, 1993), published reports studying the effect of interaural place disparity on ITD sensitivity are few (van Hoesel and Clark, 1997; Long *et al.*, 2003; Wilson *et al.*, 2003; van Hoesel, 2004) and include data from a total of only five subjects. The authors report results for an additional four subjects and relate them to the impact of interaural carrier-

TABLE I. Subject characteristics.

Subject	Etiology	Age at onset of hearing loss (years)		Age at onset of hearing-aid use (years)		Age at profound loss (years)		Age at onset of CI use (years)		Duration[a] of bilateral deprivation (years)
		R	L	R	L	R	L	R	L	
C105	Idiopathic	65	74	N/A	N/A	65	74	76	74	11
C109	Genetic	30	30	37	34	44	44	48	50	6
C120	Autoimmune	34	34	36	36	39	39	40	43	4
C128	Genetic	11	11	17	17	25	25	36	39	14

[a]Age at onset of CI use for second sound processor minus the age at profound loss for the earlier of the two ears.

frequency disparity on the modulator ITD JND measured in normal-hearing listeners and to the effect of interaural place disparity on the bilateral interaction component of the electrically evoked auditory brainstem response in cats.

II. METHODS

The Institutional Review Boards of the Massachusetts Eye and Ear Infirmary and the Massachusetts Institute of Technology reviewed and approved the methods used in the studies reported here.

A. Subjects, implants, and sound processors

Results are reported from four subjects with bilateral Advanced Bionics CII CIs who have been followed for a period of 2–4 years. They are part of an ongoing effort to understand how basic ITD and ILD sensitivity limits functional capabilities such as lateralization and speech reception in the presence of a spatially separated interferer. Our approach has been to more completely characterize ITD and ILD sensitivity using a wider range of stimuli and conditions for each of a smaller number of subjects than is practical for larger subject populations. The results reported here are extracted from this accumulating data set and are not from single studies, each designed to address the specific issues discussed in this report.

Table I lists the etiologies of the subjects and the age of each subject at the onset of: hearing loss, hearing-aid use, profound hearing impairment, and CI use. Subject C105's idiopathic loss was sudden onset (never wore hearing aids), and C120's autoimmune loss was rapidly progressive. All subjects were postlingually deafened, and all reports are consistent with our presumption that they had normal binaural hearing before the onset of their hearing loss.

The two ears of each subject were implanted sequentially with the first-implanted cochlea receiving an Advanced Bionics HiFocus electrode array with a "positioner" designed to place the array in a modiolar position. The second-implanted ears also received HiFocus electrode arrays: C105 and C109 with a positioner and C120 and C128 without a positioner. The age at which each ear began receiving daily stimulation is given in Table I.

The authors partition the subjects' listening experience with CIs into three periods that were part of the original study design and to which the subjects consented prior to implantation of the second device. Period I began with the implantation of their first ear, lasted for at least 18 months, and ended with the implantation of their second ear. During this monolateral listening period, the authors assume that the subjects developed monolateral listening strategies for speech reception (in quiet and in noise) and localization.

Period II began with implantation of their second implant. During this listening period, the subjects continued *monolateral* use of their first implant system for at least 8 months (C105: 347 days, C109: 272 days, C120: 329 days, and C128: 271 days). The only time their second implant produced stimulation was during laboratory testing sessions designed to (1) identify interaural electrodes matched in cochleotopic position to be paired in bilateral sound-processing strategies and (2) evaluate localization and speech-reception capabilities using monolateral and bilateral sound processors during a period when they presumably continued to employ a monolateral listening strategy in their day-to-day monolateral implant system use.

Listening period III began at least 8 months after implantation of the second ear when the second sound processor was fitted for daily use. Since then, subjects have used two sound processors that implement a version of the continuous interleaved sampling (CIS) strategy (Wilson *et al.*, 1991). Because the two sound processors are not synchronized, the ITD between the carrier pulses generated by the output signals of the two processors is not controlled and can drift with time. The envelope ITD between the two processors accurately reflects the ITD at the subject's microphones.

Laboratory testing sessions during listening period III were designed to (1) characterize ITD and ILD sensitivity and (2) evaluate localization and speech-reception capabilities using monolateral and bilateral sound processors during a period when subjects would be expected to use the bilateral listening strategy they develop during day-to-day use of their bilateral implant system use. The results presented in this report focus on measures of ITD sensitivity made using single-interaural electrode pairs stimulated with pulse trains during listening periods II and III.

The processor controlling the first implant in each subject mapped CIS analysis channels 1 (lowest frequency band) through 16 (highest frequency band) to electrodes 1 (most apical) through 16 (most basal). After implantation of the second implant and during listening period II, a combination of measures (e.g., binaural fusion, interaural pitch ranking, and interaural time sensitivity) was used to identify spatially matched electrodes across the two ears and to guide

TABLE II. Devices and performance.

Subject	Implant model (R and L)	Electrode model (R and L)	Electrode positioner R	Electrode positioner L	Strategy (R and L)	Channel-electrode map (Table III)	Pulse rate (pps)	Pulse width (μs)	Latest NU-6 word score R (%)	Latest NU-6 word score L (%)
C105	CII	HiFocus	Yes	Yes	CIS-16	2	1450	21.6	28	38
C109	CII	HiFocus	Yes	Yes	CIS-16	1	1450	21.6	94	86
C120	CII	HiFocus	Yes	No	CIS-16	1 and 3	2320	13.5	84	70
C128	CII	HiFocus	Yes	No	CIS-16	1 and 4	2320	13.5	84	86

the mapping of analysis channel to electrode in the second implant system (Eddington *et al.*, 2002; Eddington *et al.*, 2003). Table II summarizes the implants, the sound-processing strategies, and the latest monolateral speech-reception performance. The column "Channel-electrode map" identifies the channel-to-electrode map used for each subject. These maps are specified in Table III.

The preliminary measures used to evaluate interaural electrode-array offset suggested minor differences for subjects C105, C109, and C120. The fusion and pitch ranking measures for C128 indicated an interaural electrode-array offset of approximately two electrodes for the most apical electrodes, tapering to a single-electrode offset for the more basal electrodes. In the cases of C120 and C128, two programs were downloaded to the second sound processor: one to implement channel-to-electrode map 1 and another to implement either map 3 (C120) or map 4 (C128). C120 and C128 have been switching between the two programs approximately every two days since they began wearing two sound processors. They both report not being able to distinguish between the two maps, and the authors are not able to measure significant differences in localization or speech reception in quiet or noise between the two programs.

The results reported below were measured using single interaural electrode pairs. Whether or not these interaural pairs were also paired by analysis channel in the channel-to-electrode map(s) used by the subject in listening period III is documented in the text and figure captions.

B. Stimulus generation for psychophysical studies

ITD sensitivity was measured using single pulses and fixed-amplitude pulse trains. All pulses were biphasic (typically 27 μs/phase [1]). Stimuli were delivered monopolarly (with the receiver/stimulator cases serving as extracochlear return electrodes) to two intracochlear electrodes, one in each cochlea (an interaural pair). Custom software and a bilateral Clarion Research Interface (CRI2 manufactured by Advanced Bionics Corporation) were used to control the subjects' bilaterally implanted receivers/stimulators and to synchronize stimuli across ears to within 1 μs. ITD resolution of the controller was 13.5 μs. ITDs were generated by

TABLE III. Mapping of channels to electrodes.

Analysis channel	Map 1 Left EL	Map 1 Right EL	Map 2 Left EL	Map 2 Right EL	Map 3 Left EL	Map 3 Right EL	Map 4 Left EL	Map 4 Right EL
1	1	1	1	1[a]	2	1	1[b]	1
2	2	2	2	1[a]	3	2	1[b]	2
3	3	3	3	2	4	3	1[b]	3
4	4	4	4	3	5	4	2	4
5	5	5	5	4	6	5	3	5
6	6	6	6	5	7	6	4	6
7	7	7	7	6	8	7	5	7
8	8	8	8	7	9	8	6	8
9	9	9	9	8	10	9	7,8	9
10	10	10	10	9	11	10	9	10
11	11	11	11	10	12	11	10	11
12	12	12	12	11	13	12	11	12
13	13	13	13	12	14	13	12	13
14	14	14	14	13	15	14	13	14
15	15	15	15	14	16[c]	15	14	15
16	16	16	16	15,16	16[c]	16	15	16

[a]These electrodes were driven by a custom analysis channel with the combined bandwidth of the map 1 analysis channels 1 and 2.
[b]These electrodes were driven by a custom analysis channel with the combined bandwidth of the map 1 analysis channels 1–3.
[c]These electrodes were driven by a custom analysis channel with the combined bandwidth of the map 1 analysis channels 15 and 16.

delaying the stimulus at one electrode relative to the other, producing an onset ITD and ongoing ITD that are equal.

All tests were conducted at stimulation levels adjusted to produce a comfortable sensation level in each ear using the following three-step procedure. Prior to each run, the monolateral-left stimulus amplitude was adjusted to elicit a most comfortable sensation level. The right stimulus level was then adjusted to match the left sensation level during repeated sequential left-right stimulation. Finally, the "level-matched" right/left stimuli were played simultaneously (ITD=0 μs), and, when needed, the right-ear amplitude was adjusted to center the resulting sound image. If bilateral stimulation did not elicit a single sound image (e.g., when the individual electrodes of an interaural electrode pair were widely separated in cochlear place), the level-matched amplitudes were used. The "centered" or level-matched amplitudes were considered the "zero ILD" condition even though the actual stimulation levels in each ear were often different (typically by less than 1 dB).

For each subject, interaural electrode pairs mapped to the same analysis channel in the two asynchronous sound processors used in the field are referred to as processor-paired electrodes. Electrode pairs are denoted by the subject ID followed by the left and right electrode number (e.g., C109:L3R3).

C. Adaptive lateralization test

ITD sensitivity was measured using an adaptive (two-down, one-up, 14 reversals), two-alternative forced-choice procedure that targeted the 70.7% level on the psychometric function (Levitt, 1971). The first-interval stimulus was always the reference (ITD=0 μs). The adaptively determined ITD magnitude was applied to the second-interval stimulus with the leading side randomly selected. This procedure is sometimes referred to as a center-side or reminder task, in contrast to the two-interval, two-alternative forced-choice task, and it typically estimates a JND about $\sqrt{2}$ less than a two-cue, two-interval task where an opposite ITD is applied during each interval (Hartmann and Rakerd, 1989). The subject used a keyboard to indicate whether the probe was lateralized to the left or the right of the reference position. Correct-answer feedback was given after each trial.

Informal testing at the beginning of each run estimated the ITD JND and determined a starting ITD that the subject could easily lateralize. In cases where the subject was not able to perform above chance at the procedure's maximum ITD (2 ms), the JND could not be measured (CNM in Fig. 1). A starting step size was chosen and then was reduced by half after the first peak reversal and further reduced (by half again) after the second peak reversal in the adaptive track. Based on the informal testing, the starting step size was typically set at 108 or 54 μs (resulting in a final step size of 27 or 13.5 μs for stimuli associated with the last ten reversals of the adaptive track) unless the informal testing revealed very poor ITD sensitivity, in which case a larger starting step size was sometimes used. After about 80% of the data reported here had been collected, the authors switched to a constant scaling factor of 2 dB (multiplying or dividing by a factor of 1.26 since $10^{(2/20)} \sim 1.26$) for the adaptive runs (Saberi, 1995).

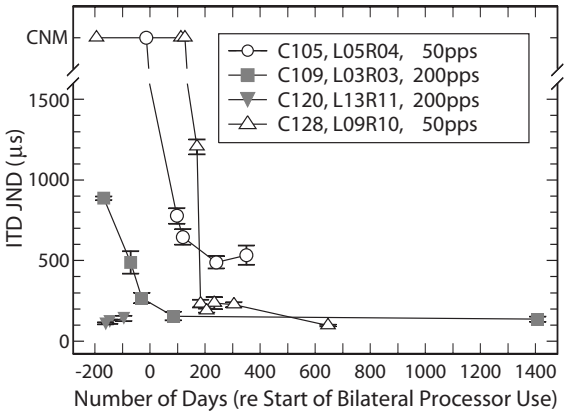


FIG. 1. ITD JND plotted as a function of time (relative to the day each subject began wearing their second sound processor). Stimuli were constant-amplitude, biphasic pulse trains (300-ms duration) with the repetition rate, subject, and interaural electrode pair identified in the legend. Symbols located above the axis break mark days when an ITD JND could not be measured (CNM) because the subject could not distinguish zero ITD from ITDs greater than 1500 μs or because stimulation of the interaural pair elicited multiple (non-fused) sound images. In cases where multiple measures were made using the same stimulus waveform on the same day, those runs were combined to determine the JND (see Sec. II). Error bars represent the standard error of these multiple measurements.

A psychometric function was estimated from the run data using a generalized linear model (binomial distribution function and probit link function). The JND was defined as the ITD corresponding to a 70.7% correct response rate. In cases where multiple runs were conducted for the same stimulus configuration and parameter set, the trials from all runs were combined and the psychometric function estimated from the combined data set. The combined set of the last eight reversals of each run was used to compute an estimate of the standard error.

Measures of ILD sensitivity used the same methods as those used to measure the ITD JND. The adaptively determined ILD was applied to the second-interval stimulus, and the side incremented in level was randomly selected. Starting ILDs were selected that elicited easily lateralized sound images (typically 2 dB), and the beginning step size was 0.4 dB.

III. RESULTS

A. ITD sensitivity as a function of time

Figure 1 plots ITD JNDs measured with unmodulated pulse trains as a function of time for the four example subject/stimulus-parameter combinations identified in the legend. Time is referenced to the day each subject began wearing his/her second sound processor. As noted in Sec. II, during the period between implantation of the second ear and the daily use of their second processor (i.e., before day 0 in the figure), the subjects continued monolateral listening in the field with the same sound processor used before their

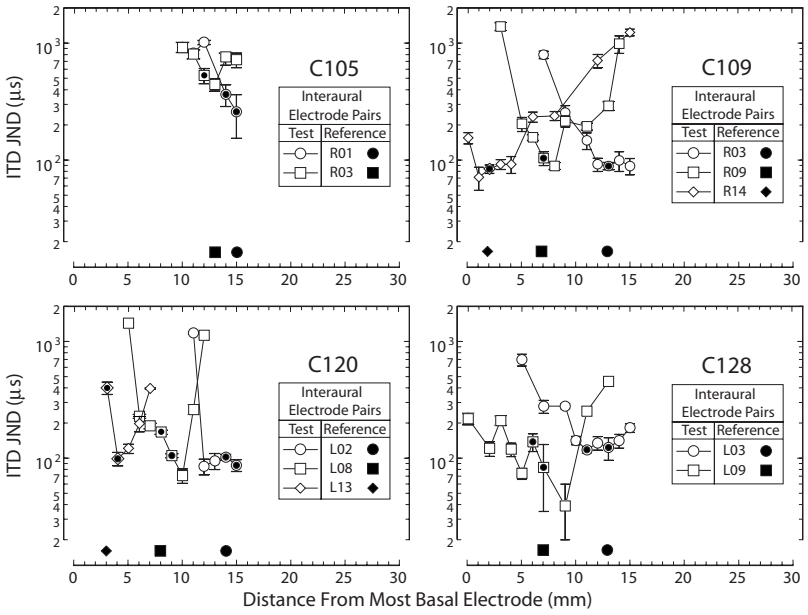


FIG. 2. ITD JND plotted as a function of the position of the test electrode (open symbols) paired with the contralateral reference electrodes (filled symbols near the abscissa) identified in each legend. Each panel shows data from one subject as identified above the legend. Both the test and reference electrode arrays consisted of 16 electrodes (~1-mm spacing, numbered from most apical to most basal). Reference electrodes are identified by side (R=right; L=left) and electrode number in the legend. Position of the reference electrode in its electrode array is shown by the filled symbols along the bottom of each panel. Open symbols represent test electrodes that are paired with the reference electrode identified by the same filled symbol. Open symbols enclosing a filled circle mark the test electrode(s) paired with the corresponding reference electrode in the subject's sound-processing system. Because subjects C120 and C128 alternated between two channel-to-electrode maps (see Table III), two test electrodes are marked for each reference electrode. Two test electrodes are also marked for subject C105 to identify the two left electrodes paired with electrode R01 in the subject's sound-processing system (see map 2 of Table III). Stimuli were 300-ms pulse trains at 50 pps for C105, C109, and C128 and at 200 pps for C120.

second surgery. The only time they experienced bilateral stimulation in this period was during laboratory testing sessions.

The data represented by gray-filled symbols are examples from subjects C109 and C120 showing that their sensitivity to ITD was apparent early in their testing and before daily use of their second sound processor The interaural pairs selected for plotting are among the first demonstrating ITD sensitivity. The JNDs were measured using the 200-pps rate of the stimulus used to search for ITD sensitivity. These two subjects also demonstrated early ILD sensitivity with JNDs of 0.22 dB for C109 and 0.28 dB for C120 by 180 days before beginning bilateral sound-processor use (day −180).

In the case of C105 and C128 (open symbols), ITD sensitivity did not develop until well into their daily experience with bilateral CI listening. By this time, the authors had added a search stimulus of 50 pps because of its greater ITD sensitivity. They expected that the emergence of ITD sensitivity would have occurred at an even later point in time using the 200-pps search stimulus because ITD JNDs could often not be measured at this higher rate when sensitivity was first detected using the 50-pps search stimulus. For instance, at day +98, an ITD JND of 777 μs was measured for C105 using the 50-pps stimulus but could not be measured (JND$>$1500 μs) using the 200-pps stimulus.

Like subjects C109 and C120, subjects C105 and C128 demonstrated relatively early ILD sensitivity with ILD JNDs measured before day −130 of 0.50 dB for C105 and 0.18 dB for C128. These results indicate that it is unlikely that the measurement of early ITD sensitivity for subjects C105 and C128 was limited by the subjects' ability to perform the lateralization task. The early ILD-based lateralization does not demonstrate that binaural processing for ILD cues developed before ITD cues because the authors cannot rule out the use of monaural level cues.

The ITD JND measured in subjects C105, C109, and C128 generally improved for several months after ITD sensitivity was first observed. The improvements in C109's and C128's sensitivity are dramatic with ITD JNDs below 200 μs while C105's is more modest with best JNDs of approximately 500 μs. C120's sensitivity started out better than that C105 could achieve after more than a year of bilateral listening experience. The results marked by gray-filled squares illustrate the general trend for C109 of ITD sensitivity improving with time before she began using her second sound processor even though her exposure to bilateral stimuli consisted of only 2- or 3-hour testing sessions once or twice each month during this period. The data presented in subsequent sections of this report were measured after the ITD sensitivity had stabilized.

B. ITD JND as a function of interaural electrode separation

Figure 2 plots measures of ITD JND made in four subjects as a function of the contralateral test electrode used to form a single-interaural pair with the reference electrodes identified in the legends. Stimuli were 300-ms pulse trains at 50 pps for C105, C109, and C128 and at 200 pps for C120 (because C120 was lost to study before the 50-pps measures could be made). In the case of C105, only measures for apical references were made because the ITD JNDs for the most sensitive interaural pairs of the more basally placed reference electrodes (distance from most basal electrode less than 8 mm) were greater than 700 μs, and when the test electrode was changed from this most sensitive position, ITD JND was typically not measurable. Measures using a basal reference were not made in C128 because of time constraints.

The electrode combination showing the greatest ITD sensitivity for a specific reference was often not the pair judged most similar in cochleotopic position. The symbols enclosing filled circles mark test electrodes that were paired with the specified reference electrode in the subject's sound processors based on the combined measures described in

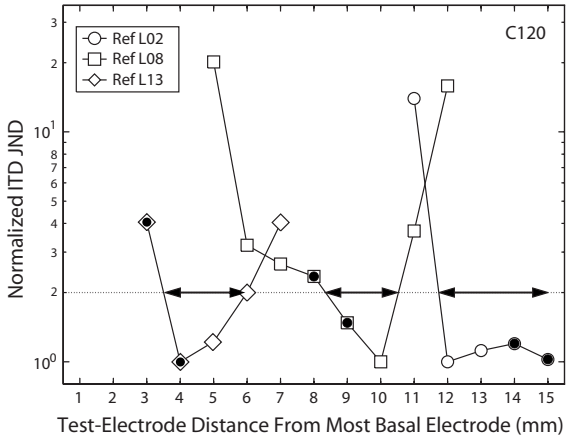


FIG. 3. Normalized ITD JND for three reference electrodes (see legend) plotted as a function of the contralateral test electrode position for subject C120 to illustrate the half-width metric. The horizontal dotted line marks the normalized ITD JND that is a factor of 2 greater than the minimum. The horizontal arrows identify the ranges (half-widths) defined by the apical end of the electrode array and the intersections of the dotted line with the line segments associated with each reference condition.

Sec. II. In only four of the ten reference conditions tested did the processor-paired electrodes correspond to the interaural pair with lowest ITD JND.

In order to compare the degree to which test electrode influences ITD sensitivity across reference conditions and subjects, the authors normalized the measures associated with each reference condition by the minimum ITD JND for that condition and, as illustrated in Fig. 3, determined the range (half-width) over the length of the electrode array within which the ITD JND is expected to be within a factor of 2 of the minimum for each reference electrode.

The half-widths computed from the Fig. 3 data are listed in Table IV and range from 2.1 to 5.7 mm. The ANOVA-computed effects of subject and electrode position (apical, middle, or basal) on half-width were not significant

IV. DISCUSSION

Like most studies of ITD sensitivity in CI users, our measures of ITD JND show a substantial range across subjects. Figures 1 and 2 illustrate the poorer ITD sensitivity of C105 compared to the other subjects. C105 is also the oldest of the subjects (Table I), and her speech-reception performance is modest with a maximum NU6 word score of 38% correct compared to the others who scored at least 84%.

TABLE IV. ITD JND half-widths.

Subject	Reference electrode	ITD JND half-width (mm)
C105	R01	0.8
C105	R03	4.6
C109	R03	4.7
C109	R09	3.3
C109	R14	4.8
C120	L02	3.3
C120	L08	2.1
C120	L13	2.5
C128	L03	5.7
C128	L09	2.5

TABLE V. Deprivation and bilateral listening experience. Three metrics are listed for each subject: (1) the duration of bilateral deprivation in years (see Table I for definition), (2) the first day (referenced to when each subject began bilateral CI use) that ITD JNDs less than 700 μs were measured, and (3) estimated bilateral listening experience each subject experienced up to the day when ITD JNDs less than 700 μs were first measured. Bilateral listening experience before start of bilateral CI use was restricted to laboratory testing sessions. Bilateral listening experience after the start of bilateral CI use was estimated to be 14 h/day.

Subject	Duration of bilateral deprivation (years)	Day when ITD JND $<$ 700 μs (re start of bilateral CI use)	Bilateral listening experience when ITD JND $<$ 700 μs (estimated hours)
C120	4	−161	8
C109	6	−70	14
C105	11	119	1666
C128	14	184	2576

A. ITD sensitivity as a function of time

The results plotted in Fig. 1 show that the point at which subjects demonstrate stable ITD sensitivity varies greatly across subjects. Three of the four subjects showed substantial improvements in ITD sensitivity over time frames of 6 months or more. Two of these three subjects did not begin improving until after the onset of daily bilateral stimulation. Only C120 showed relatively good ITD sensitivity directly after implantation. Another example of early sensitivity is subject CI3 of Laback *et al.* (2007) who suffered a symmetric, bilateral deafness for 2 months before implantation. After only 1 month of bilateral implant listening experience, an ITD JND of 30 μs was measured using four-pulse, 100-pps unmodulated pulse trains.

The right column of Table I shows that subjects C109 and C120, who demonstrated ITD sensitivity before beginning daily bilateral CI use, experienced relatively short durations of bilateral deprivation (6 and 4 years, respectively). The two subjects experiencing longer periods of deprivation (11 years for C105 and 14 years for C128) showed minimal improvement in testing before day-to-day binaural listening was available, and they required considerable bilateral listening experience before ITD sensitivity could be measured. Table V shows the subjects ordered by duration of bilateral deprivation and lists two metrics: (1) the day (day 700) when ITD JNDs below 700 μs were first measured (this day was referenced to the day each subject began wearing his/her second sound processor) and the estimated hours of bilateral CI listening experienced by each subject before day 700. The listening experience metric splits the subjects into two widely separated groups: those exhibiting ITD JNDs less than 700 μs before and after beginning daily use of bilateral sound processors. While the correspondence of this delayed emergence of ITD sensitivity with the duration of bilateral deprivation in the small number of subjects studied does not establish an association, it does suggest the hypothesis that bilateral deprivation influences the emergence of bilateral sensitivity in CI subjects. Such a hypothesis is consistent

with the deafness-induced plasticity observed in the structural and functional changes in auditory brainstem and midbrain reported in animals (reviews: Shepherd and Hardie, 2001; Hartmann and Kral, 2004; Moore and King, 2004) and warrants further study.

The variability in the emergence of ITD sensitivity and the amount of bilateral listening experience some subjects require before ITD sensitivity stabilizes complicates the comparison of across-subject results when some subjects are studied within the first 6 months of bilateral implantation and others after significantly longer use. If within-subject ITD-sensitivity measures for different stimulus parameters/conditions are also distributed within and beyond the first 6 months of bilateral sound-processor use, the comparison of conditions within and across subjects can be problematic.

B. ITD sensitivity as a function of electrode separation

In order to compare the half-widths computed for our subjects (Table IV) with those for subjects from other studies, the authors reviewed the published and unpublished reports that used unmodulated pulse trains to measure ITD JNDs for a reference electrode (the electrode of an interaural pair that is held constant) as a function of the location of the contralateral test electrode. The authors identified the measures from 12 reference electrodes in seven subjects that are plotted in Fig. 4 (van Hoesel and Clark, 1997; Lawson *et al.*, 2002; Wolford *et al.*, 2003; van Hoesel, 2004). Even though half of the reference conditions tested were in P1 and P2 who exhibited very poor ITD sensitivity (ITD JNDs >600 μs) and substantial interaural offset (>4.5 mm) of their electrode arrays (see van Hoesel and Clark, 1997), this body of data is consistent with the results of Fig. 2 in suggesting that the test electrode paired with the reference can make a substantial difference in the pair's sensitivity to ITD.

In Figs. 2 and 4, the electrode combination showing the greatest ITD sensitivity for a specific reference was often not the pair judged most similar in cochleotopic position. The symbols enclosing filled circles (dark dots) in Figs. 2 and 4 mark the test electrode that elicited a pitch sensation matching that elicited by stimulation of the reference electrode. In only 10 of the 22 reference conditions of the combined results did the pitch-matched pair exhibit the smallest ITD JND.

The distribution of the half-width for the combined Figs. 2 and 4 data sets suggested that the half-widths computed for subjects P1 and P2 were larger (mean: 6.3 mm; range: 4.0–7.9 mm) than the others (mean: 3.7 mm; range: 0.8–6.75 mm). A t-test confirmed the significance of this difference ($t=-3.1$, df$=18$, and $p<0.001$). When results from P1 and P2 were excluded from the combined results, ANOVA found only nonsignificant effects on the half-widths computed from the results of Figs. 2 and 4 for the factors subject, electrode type, and electrode-pair position along the electrode array (apical, middle, or basal).

The results shown in Figs. 2 and 4 illustrate the impact of spatial disparity on ITD sensitivity. Given that the tonotopic organization of the auditory system includes neurons of

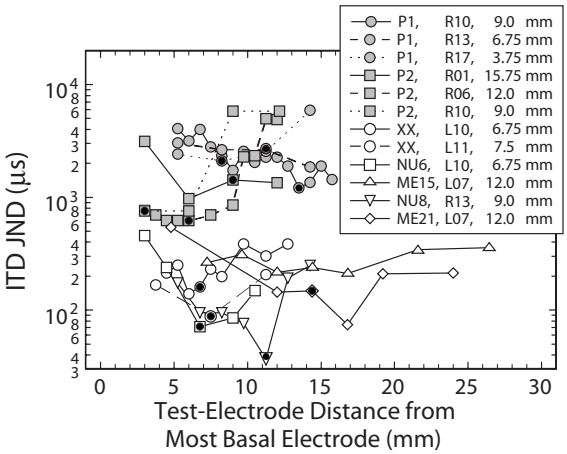


FIG. 4. ITD JND measured using single interaural electrode pairs and plotted as a function of position of the contralateral test electrode paired with the reference electrodes identified in the legend (R=right and L=left). The legend also specifies the subject and position (distance from most basal electrode of the reference electrode array) for each reference electrode. Test electrode position is given in distance from the most basal electrode of its electrode array to facilitate comparison between different arrays: Cochlear Corp. (22 electrodes, 0.75-mm spacing; subjects P1 and P2 with apical-to-basal numbering and subjects XX, NU6, and NU8 with basal-to-apical numbering) and Med-El Corp. (12 electrodes, ~2.4-mm spacing; subjects ME15 and ME21). Monopolar electrode configurations were used in all subjects except for P1 and P2 where electrode numbers identify the more apical of the BP+1 bipolar pair. Symbols enclosing a filled circle mark test electrodes reported to be place matched with their respective reference electrode based on a pitch criterion. In the case of ME15, a place match could not be obtained: reference electrode L07 was judged lower in pitch than R07 and higher than R06 (Lawson *et al.*, 2002). It is not always clear whether the bilateral sound processors used by these subjects paired the place-matched electrodes by analysis channel. Stimuli were constant-amplitude pulse trains with repetition rates varying from 50 to 200 pps depending on the study. Data taken from the original publications/reports: NU6 and ME15 (Lawson *et al.*, 2002); P1 and P2 (van Hoesel and Clark, 1997); XX (van Hoesel, 2004); NU8 and ME21 (Wolford *et al.*, 2003).

the brainstem (e.g., Guinan *et al.*, 1972; Yin and Chan, 1990) and midbrain (e.g., Kuwada *et al.*, 1984), it is not surprising that ITD sensitivity decreases as the spatial disparity between stimulation electrodes of an interaural pair increases. The impact of this organization was also seen by Smith and Delgutte (2007) in their measures of the binaural interaction component (BIC) of electrically evoked auditory brainstem responses in cats. Based on their mean BIC data plotted as a function of interaural electrode offset [their Fig. 4(B)], the authors estimated the BIC half-width to be approximately 4.5 mm, well within the range of ITD JND half-widths found in the human CI subjects. The ITD JND and BIC half-widths are also consistent with the 3–5 mm spatial spread of excitation estimated by Long *et al.* (2003) from the monaural forward masking results reported by Cohen *et al.* (2001).

Nuetzel and Hafter (1981) measured the impact of interaural carrier-frequency disparity on the modulator ITD JND measured in normal-hearing listeners using sinusoidally amplitude-modulated high-frequency tones. The authors transformed their Figs. 2 and 3 results (ITD JND as a function of left-ear carrier frequency for a reference right-ear carrier) into ITD JND as a function of offset in cochlear place using Liberman's (1982) cochlear frequency map adjusted for humans (Greenwood, 1990). The mean ITD JND half-width computed from these acoustically measured data

was 0.7 mm (range: 0.4–1.1 mm). This is on the order of a critical band and significantly smaller than the ITD JND half-width (mean: 3.7 mm) the authors measured in implantees ($t=5.0$, df=18, and $p<0.001$).

In every day life, CI users will experience ITD information distributed across multiple sound-processing channels because the natural listening environment often includes multiple broadband sound sources. The ability of CI systems to elicit patterns of spike activity that accurately represent such environments is likely to be compromised by half-widths for electric stimulation that the authors estimate to be five times greater than those for normal hearing.

C. Selecting interaural pairs for bilateral sound processors

One challenge facing clinicians configuring sound processors for bilateral implantees is establishing a mapping of filter-bank channels to interaural electrodes that ensures that the ITD of a within-band signal is presented to a pair of electrodes that is most sensitive to the ITD. Our results demonstrate that some implantees exhibit ITD sensitivity at the time of fitting. In these cases, using direct measures of ITD sensitivity to guide the pairing of interaural electrodes is possible. This technique will probably provide the best chance of a functional binaural benefit. Our results also show that some subjects do not have sensitivity at the time of fitting, forcing clinicians to rely on more indirect techniques.

Interaural pitch comparisons are sometimes used to estimate relative cochleotopic position to aid in the interaural pairing of electrodes. Our data and others (e.g., Long *et al.*, 2001) show that this technique does not guarantee identification of electrode pairs with optimal ITD sensitivity. Typical current practice is to ignore the issue and program each sound processor as if the implants were monolateral. The relatively large half-widths (mean: 3.7 mm) described in this study indicate that close but imperfect matching, while not optimum, may retain some useful ITD sensitivity when using these approaches.

Pelizzone *et al.* (1990) demonstrated that the BIC can be recorded in a bilaterally implanted human subject and, as noted in the previous section, the results of Smith and Delgutte (2007) measured in cats demonstrate that the BIC is sensitive to interaural electrode offset. If the BIC is present before ITD sensitivity can be measured psychophysically, it would be a valuable tool for pairing interaural electrodes and invites further investigation.

ACKNOWLEDGMENTS

We are especially indebted to the research subjects for their generous contribution of time and effort. We thank Bertrand Delgutte and two anonymous reviewers for valuable comments on the manuscript and recognize the support of Advanced Bionics Corp. for their donation of the second-ear implant systems. NIH Grant Nos. R01 DC005775 and R01 DC007528 provided major support. B.B.P. was partially supported by NIH Grant No. T32 DC00038.

[1]In order to reach comfortable listening levels at reasonable stimulus amplitudes using single-pulse stimuli, the phase duration used in the psychophysical experiments was longer than the 13.5 and 21.6 μs phase durations used by the subjects' sound-processing systems (see Table III).

Arsenault, M. D., and Punch, J. L. (**1999**). "Nonsense-syllable recognition in noise using monaural and binaural listening strategies," J. Acoust. Soc. Am. **105**, 1821–1830.

Bronkhorst, A. W., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508–1516.

Cohen, L. T., Saunders, E., and Clark, G. M. (**2001**). "Psycohphysics of a prototype peri-modiolar cochlear implant electrode array," Hear. Res. **155**, 63–81.

Eddington, D. K., Poon, B. B., Colburn, H. S., Noel, V., Herrmann, B., Tierney, J., and Whearty, M. (**2003**). "Speech processors for auditory prostheses: Fifth quarterly progress report," Massachusetts Institute of Technology, Cambridge.

Eddington, D. K., Tierney, J., Noel, V., Herrmann, B., Whearty, M., and Finley, C. C. (**2002**). "Speech processors for auditory prostheses: Third quarterly progress report," Massachusetts Institute of Technology, Cambridge.

Gordon, K. A., Valero, J., and Papsin, B. C. (**2007**). "Auditory brainstem activity in children with 9–30 months of bilateral cochlear implant use," Hear. Res. **233**, 97–107.

Grantham, D. W., Ashmead, D. H., Ricketts, T. A., Haynes, D. S., and Labadie, R. F. (**2008**). "Interaural time and level difference thresholds for acoustically presented signals in post-lingually deafened adults fitted with bilateral cochlear implants using CIS+processing," Ear Hear. **29**, 33–44.

Grantham, D. W., Ashmead, D. H., Ricketts, T. A., Labadie, R. F., and Haynes, D. S. (**2007**). "Horizontal-plane localization of noise and speech signals by postlingually deafened adults fitted with bilateral cochlear implants," Ear Hear. **28**, 524–541.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Guinan, J. J., Norris, B. E., and Guinan, S. S. (**1972**). "Single auditory units in the superior olivary complex II: Locations of unit categories and tonotopic organization," Int. J. Neurosci. **4**, 147–166.

Hafter, E. R., and Dye, R. H., Jr. (**1983**). "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.

Hartmann, R., and Kral, A. (**2004**). "Central responses to electrical stimulation," in *Cochlear Implants: Auditory Prostheses and Electric Hearing*, edited by F.-G. Zeng, R. R. Fay, and A. N. Popper (Springer-Verlag, New York), pp. 213–285.

Hartmann, W. M., and Rakerd, B. (**1989**). "On the minimum audible angle—A decision theory approach," J. Acoust. Soc. Am. **85**, 2031–2041.

Klumpp, R. G., and Eady, E. R. (**1956**). "Some measurements of interaural time difference thresholds," J. Acoust. Soc. Am. **28**, 859–860.

Kuwada, S., Yin, T. C. T., Syka, J., Buunen, T. J. F., and Wickesberg, R. E. (**1984**). "Binaural interaction in low-frequency neurons in inferior colliculus of the cat. IV. Comparison of monaural and binaural response properties," J. Neurophysiol. **51**, 1306–1325.

Laback, B., Majdak, P., and Baumgartner, W.-D. (**2007**). "Lateralization discrimination of interaural time delays in four-pulse sequences in electric and acoustic hearing," J. Acoust. Soc. Am. **121**, 2182–2191.

Laback, B., Pok, S. M., Baumgartner, W. D., Deutsch, W. A., and Schmid, K. (**2004**). "Sensitivity to interaural level and envelope time differences of two bilateral cochlear implant listeners using clinical sound processors," Ear Hear. **25**, 488–500.

Lawson, D. T., Wilson, B. S., Zerbi, M., van den Honert, C., Finley, C. C., Farmer, J. C., Jr., McElveen, J. T., Jr., and Roush, P. A. (**1998**). "Bilateral cochlear implants controlled by a single speech processor," Am. J. Otol. **19**, 758–761.

Lawson, D. T., Wolford, R., Wilson, B. S., and Schatzer, R. (**2002**). "Speech processors for auditory prostheses: First quarterly progress report," Research Triangle Institute, Research Triangle Park.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Liberman, M. C. (**1982**). "The cochlear frequency map for the cat: Labeling auditory-nerve fibers of known characteristic frequency," J. Acoust. Soc. Am. **72**, 1441–1449.

Litovsky, R. Y., Johnstone, P. M., Godar, S., Agrawal, S., Parkinson, A., Peters, R., and Lake, J. (**2006b**). "Bilateral cochlear implants in children:

Localization acuity measured with minimum audible angle," Ear Hear. **27**, 43–59.
Litovsky, R., Parkinson, A., Arcaroli, J., and Sammeth, C. (**2006a**). "Simultaneous bilateral cochlear implantation in adults: A multicenter clinical study," Ear Hear. **27**, 714–731.
Long, C. J., Eddington, D. K., Colburn, H. S., and Rabinowitz, W. M. (**2001**). "Sensitivity to interaural time difference as a function of interaural electrode position in a cochlear implant user," in 2001 Conference on Implantable Auditory Prostheses, Pacific Grove, CA, p. 20.
Long, C. J., Eddington, D. K., Colburn, H. S., and Rabinowitz, W. M. (**2003**). "Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user," J. Acoust. Soc. Am. **114**, 1565–1574.
Majdak, P., Laback, B., and Baumgartner, W. D. (**2006**). "Effects of interaural time differences in fine structure and envelope on lateral discrimination in electric hearing," J. Acoust. Soc. Am. **120**, 2190–2201.
Moore, D. R., and King, A. J. (**2004**). "Plasticity of binaural systems," in *Plasticity of the Auditory System*, edited by T. N. Parks, E. W. Rubel, R. R. Fay, and A. N. Popper (Springer-Verlag, New York), pp. 96–172.
Nopp, P., Schleich, P., and D'Haese, P. (**2004**). "Sound localization in bilateral users of MED-EL COMBI 40/40+ cochlear implants," Ear Hear. **25**, 205–214.
Nuetzel, J. M., and Hafter, E. R. (**1981**). "Discrimination of interaural delays in complex waveforms: Spectral effects," J. Acoust. Soc. Am. **69**, 1112–1118.
Oldfield, S. R., and Parker, S. P. A. (**1986**). "Acuity of sound localisation: A topography of auditory space. III. Monaural hearing conditions," Perception **15**, 67–81.
Pelizzone, M., Cosendai, G., and Tinembart, J. (**1999**). "Within-patient longitudinal speech reception measures with continuous interleaved sampling processors for ineraid implanted subjects," Ear Hear. **20**, 228–237.
Pelizzone, M., Kasper, A., and Montandon, P. (**1990**). "Binaural interaction in a cochlear implant patient," Hear. Res. **48**, 287–290.
Peters, B. R., Litovsky, R., Parkinson, A., and Lake, J. (**2007**). "Importance of age and postimplantation experience on speech perception measures in children with sequential bilateral cochlear implants," Otol. Neurotol. **28**, 649–657.
Poon, B. B. (**2006**). "Sound localization and interaural time sensitivity with bilateral cochlear implants," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, p. 170.
Ricketts, T. A., Grantham, D. W., Ashmead, D. H., Haynes, D. S., and Labadie, R. F. (**2006**). "Speech recognition for unilateral and bilateral cochlear implant modes in the presence of uncorrelated noise sources," Ear Hear. **27**, 763–773.
Saberi, K. (**1995**). "Some considerations on the use of adaptive methods for estimating interaural-delay thresholds," J. Acoust. Soc. Am. **98**, 1803–1806.
Schleich, P., Nopp, P., and D'Haese, P. (**2004**). "Head shadow, squelch, and summation effects in bilateral users of the MED-EL COMBIN 40/40+ cochlear implant," Ear Hear. **25**, 197–204.
Seeber, B. U., and Fastl, H. (**2008**). "Localization cues with bilateral cochlear implants," J. Acoust. Soc. Am. **123**, 1030–1042.
Sharma, A., Dorman, M. F., and Kra, A. (**2005**). "The influence of a sensitive period on central auditory development in children with unilateral and bilateral cochlear implants," Hear. Res. **203**, 134–143.
Shepherd, R. K., and Hardie, N. A. (**2001**). "Deafness-induced changes in the auditory pathway: Implications for cochlear implants," Audiol. Neuro-Otol. **6**, 305–318.
Smith, Z. M., and Delgutte, B. (**2007**). "Using evoked potential to match interaural electrode pairs with bilateral cochlear implants," J. Assoc. Res. Otolaryngol. **8**, 134–151.
Tyler, R. S., and Summerfield, A. Q. (**1996**). "Cochlear implantation: Relationships with research on auditory deprivation and acclimatization," Ear Hear. **17**, 38S–51S.
van Hoesel, R. J. M. (**2004**). "Exploring the benefits of bilateral cochlear implants," Audiol. Neuro-Otol. **9**, 234–246.
van Hoesel, R. J. M. (**2007**). "Sensitivity to binaural timing in bilateral cochlear implant users," J. Acoust. Soc. Am. **121**, 2192–2206.
van Hoesel, R. J. M., Böhm, M., Pesch, J., Vandali, A., Battmer, R. D., and Lenarz, T. (**2008**). "Binaural speech unmasking and localization in noise with bilateral cochlear implants using envelope and fine-timing based strategies," J. Acoust. Soc. Am. **123**, 2249–2263.
van Hoesel, R. J., and Clark, G. M. (**1997**). "Psychophysical studies with two binaural cochlear implant subjects," J. Acoust. Soc. Am. **102**, 495–507.
van Hoesel, R. J. M., Ramsden, R., and O'Driscoll, M. (**2002**). "Sound-direction identification, interaural time delay discrimination and speech intelligibility advantages in noise for a bilateral cochlear implant user," Ear Hear. **23**, 137–149.
van Hoesel, R. J., Tong, Y. C., Hollow, R. D., and Clark, G. M. (**1993**). "Psychophysical and speech perception studies: A case report on a binaural cochlear implant subject," J. Acoust. Soc. Am. **94**, 3178–3189.
van Hoesel, R. J. M., and Tyler, R. S. (**2003**). "Speech perception, localization, and lateralization with bilateral cochlear implants," J. Acoust. Soc. Am. **113**, 1617–1630.
Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). "Better speech recognition with cochlear implants," Nature (London) **352**, 236–238.
Wilson, B. S., Lawson, D. T., Muller, J. M., Tyler, R. S., and Kiefer, J. (**2003**). "Cochlear implants: Some likely next steps," Annu. Rev. Biomed. Eng. **5**, 207–249.
Wolford, R., Lawson, D. T., Schatzer, R., Sun, X., and Wilson, B. S. (**2003**). "Speech processors for auditory prostheses: Fourth quarterly progress report; NIH Contract N01-DC-2-1002," Research Triangle Institute, Research Triangle.
Yin, T. C. T., and Chan, J. C. K. (**1990**). "Interaural time sensitivity in medial superior olive of cat," J. Neurophysiol. **64**, 465–488.
Zurek, P. M. (**1993**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston), pp. 255–276.

Analysis of flow-structure interaction in the larynx during phonation using an immersed-boundary method

Haoxiang Luo[a)]
Department of Mechanical Engineering, Vanderbilt University, 2301 Vanderbilt Place, Nashville, Tennessee 37235

Rajat Mittal
Department of Mechanical and Aerospace Engineering, George Washington University, Washington, DC 20052

Steven A. Bielamowicz
Division of Otolaryngology, George Washington University, Washington, DC 20052

(Received 16 July 2008; revised 20 May 2009; accepted 4 June 2009)

A recently developed immersed-boundary method is used to model the flow-structure interaction associated with the human phonation. The glottal airflow is modeled as a two-dimensional incompressible flow driven by a constant subglottal pressure, and the vocal folds are modeled as a pair of three-layered, two-dimensional, viscoelastic structures. Both the fluid dynamics and viscoelasticity are solved on fixed Cartesian grids using a sharp-interface immersed boundary method. It is found that the vibration mode and frequency of the vocal fold model are associated with the eigenmodes of the structures, and that the transition of the vibration mode takes place during onset of the sustained vibration. The computed glottal waveforms of the volume flux, velocity, and pressure are reasonably realistic. The glottal flow features an unsteady jet whose direction is deflected by the large-scale vortices in the supraglottal region. A detailed analysis of the flow and vocal fold vibrations is conducted in order to gain insights into the biomechanics of phonation. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3158942]

PACS number(s): 43.70.Bk, 43.70.Gr, 43.28.Ra [AL] Pages: 816–824

I. INTRODUCTION

Flow-structure interaction (FSI) between the air expelled by the lungs and the vocal fold (VF) tissues is the essential process that generates sound. A high-fidelity model that describes the air/VF interaction could complement experimental studies, thereby helping us understand the physics of voice production. It may also eventually help assess voice related pathologies.[1] A number of mathematical models of different complexity have been developed in the past for describing the FSI during phonation. Among them, the spring-mass-damper models are frequently used to investigate various aspects of phonation, such as the chaotic motion and asymmetry in VF vibrations.[2,3] In addition to the lumped-mass approaches, models based on the continuum mechanics of either airflow or VF tissues, or both, have been developed to simulate the laryngeal dynamics. Using the finite-element method (FEM) for the structural dynamics, Berry and Titze[4] studied the free vibration modes of a brick-shaped VF model. Berry *et al.*[5] and Alipour *et al.*[6] developed a two-/three-dimensional (2D/3D) hybrid FEM model of the VFs incorporating three tissue layers and the anisotropic material properties. Coupling this model with a 2D flow solver, they qualitatively compared the eigenmodes of the VF model with the vibration modes extracted from the FSI simulations. Recently, Thomson *et al.*[7] used the 2D FEM simulations to study the energy transfer from the airflow to the VF during the FSI, and Tao and Jiang[8] combined a 3D VF model and Bernoulli's law to investigate the anterior-posterior biphonation phenomenon.

Much work has been devoted in studying the flow field close to the VFs, the gross characteristics of the flow, and the aerodynamic forces on the VF surfaces. For example, Alipour *et al.*[6,9] studied the glottal waveforms and flow separation in the glottis. Scherer *et al.*[10,11] studied the pressure within the glottis in both stationary and driven mechanical models. Rosa *et al.*[12] presented a fully 3D model in which the dynamics of the three-layer and transversely isotropic VF was coupled with an incompressible flow solver to simulate the FSI. Using the model, the authors studied the phase difference in the VF tissue deformation and the effect of the false vocal folds (FVFs) on the pressure distribution over the laryngeal surfaces. Recently, Duncan *et al.*[13] applied an immersed-boundary (IB) method to model the FSI and examined the vorticity around the glottal exit; Tao *et al.*[3] considered a 2D viscous flow and a two-mass model to study the asymmetric glottal jet and VF vibration.

The unsteady vortex motion and turbulence are essential for the broadband noise in the human voice and thus have an important effect on the voice quality. It has also been argued that the fluctuating force produced by the vortex structures has direct impact on the VF dynamics, which in turn influences the flow and sound generation.[14] Rich fluid dynamics has been reported in a few recent experimental studies. From their particle image velocimetry measurements of a pulsatile

[a)]Author to whom correspondence should be addressed. Electronic mail: haoxiang.luo@vanderbilt.edu

 0001-4966/2009/126(2)/816/9/$25.00

flow through a stationary VF model, Erath and Plesniak[15] observed the cycle-to-cycle flipping of the glottal jet from one side of the VF to the other. Triep *et al.*[16] found that the flow field downstream the VF model is highly three-dimensional, and the vortex structures in the flow have a frequency of five times higher than the fundamental frequency of the VF. In a self-oscillating physical model experiment, Neubauer *et al.*[17] observed that the axis of the jet oscillates and exhibits an interesting wavy form. In addition, they found that the flow is marked by roll-up and convection of the large-scale vortices. Physical understanding of the phenomena reported in these works is still lacking. The asymmetry in the jet is often explained by the so called Coanda effect,[3,10,15] i.e., the tendency of the jet to attach to one of the VF medial surfaces when the glottis has a divergent shape. However, it has been found that the parallel flow entering a suddenly enlarged channel through a slit without divergence is generally asymmetric at moderate Reynolds numbers due to the downstream recirculating flow.[18] Therefore, a divergent shape of the glottis does not seem to be necessary for the flow to become asymmetric. Neubauer *et al.*[17] hypothesized that the asymmetric flow observed in their experiments is induced by the vortices downstream from the VFs. This hypothesis appears to be more plausible considering that the glottal jet indeed experiences a sudden expansion after leaving the glottis and it is likely that the jet would interact with the downstream vortices. To examine the hypothesis and understand more general behavior of the vortex motions in the supraglottal area, a computational fluid dynamics (CFD) simulation that resolves the domain away from the VFs is necessary.

One of main challenges for the CFD study of phonation is that the complex/moving geometry of the larynx is difficult to address. For this, Luo *et al.*[19] recently developed an IB method to handle the complex boundaries for a class of FSI problems. In this method, both the incompressible Navier–Stokes equation governing the flow and the Navier equation governing the elasticity of the solid are solved on stationary Cartesian grids. Details of this method and its relevance to other variants of the IB methods available in literature are provided by Luo *et al.*[19]

The goal of the present paper is to extend the work of Luo *et al.*[19] and to study the mechanism of the sustained vibration and the interaction between the glottal jet and the downstream unsteady vortices. Compared to the previous VF modeling research, the present work advances the continuum-based model using an efficient and accurate IB method and captures features of the VF vibration and glottal flow with a high level of detail. In addition, the authors will focus on the connection between the flow-induced vibration and the eigenmodes of the VFs for both the transient stage and sustained vibration. Although the relationship between the sustained vibration mode and the eigenmodes has been studied before and it was found that a particular eigenmode typically dominates the sustained vibration,[5,8] the issue has not been looked at for the transient vibration. The transient process is important as it may affect the voice onset qualities as perceived by experts.[20] Furthermore, the authors will investigate the pressure distribution on the VFs and the role of

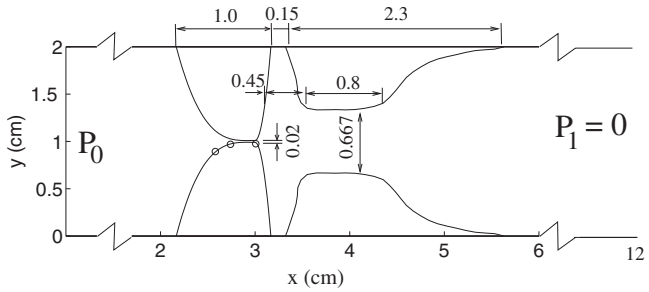


FIG. 1. The 2D computational domain (to scale) including the true VFs and FVFs where the length unit is cm. The three points marked with circles on the lower VF are chosen for analysis of the VF vibration in Sec. III.

pressure during the opening and closing phases of VF vibration. Generally speaking, the pressure does positive work on the VF during the opening phase and negative work during the closing phase. However, several factors such as the air/tissue inertia and viscosity complicate the situation. Previously, linear stability analysis[21,22] has been used to examine the issue of pressure work. In another work by Thomson *et al.*,[7] the authors found that the sign of the pressure work approximately corresponds to the glottal opening/closing. They also found that the VFs rely on their own elasticity to recover their shape during the closing phase and an auxiliary pressure force does not exist. In the present work the authors will re-examine the role of pressure during both the opening and closure of the VFs.

II. SIMULATION SET-UP AND MODELING APPROACH

The 2D computational domain representing the coronal section of the larynx located at the VF anterior-posterior midplane is shown in Fig. 1. It consists of a straight channel, a pair of deformable true VFs, and a pair of rigid FVFs downstream the true VFs. The domain is symmetric about the centerline of the channel. The geometries of the true/false VFs are chosen based nominally on the computed tomography scan data of a live person,[19] and the dimensions of the various features are described in Fig. 1. The VFs have a three-layer structure including the cover, ligament, and body,[6] as described in detail by Luo *et al.*[19]

The shortest distance between the two VFs along the medial surfaces is time-varying during the VF deformation and is defined as the glottal gap, denoted by d. In the simulations, a small threshold, $d_{\min}=0.02$ cm, is set on the glottal gap between the two VFs to prevent the flow domain from being disconnected during the VF closure. A simple contact model is incorporated to handle the VF collision. Two collision lines are set at $\pm d_{\min}/2$ off the midplane, which are non-penetrable and non-slippery for the VFs. When a point on the medial surface of one VF reaches the collision line of its side and tends to cross over, the VF is deemed to be in collision at the point. The contact point on the VF is then simply fixed on the collision line. As the VF starts to open, the kinematic constraint on the contact point is released. The VF is considered to be opening at the contact point if (1) the y velocity component at a selected point that is close to the contact point but located inside the VF points into the VF, and (2) the y force component on the VF surface at the

TABLE I. Characteristics of the volume flux waveform, including the fundamental frequency, f, duration of the open phase T_o, rising and dropping time of Q during the open phase, T_p and T_n, and the ratios $\tau_0=T_o/T$, $\tau_s=T_p/T_n$, and $q_r=Q_{\text{mean}}/Q_{\text{max}}$. The units are kPa for P_0, Hz for f, ms for T_o, T_p, and T_n, and cm^2/s for Q_{max} and Q_{mean}.

P_0	f	T_o	T_p	T_n	Q_{max}	Q_{mean}	τ_0	τ_s	q_r
0.8	162	4.74	2.55	2.19	405	180	0.77	1.16	0.444
1.0	163	4.43	2.40	2.03	545	248	0.72	1.18	0.456
1.2	164	4.43	2.47	1.97	755	314	0.73	1.26	0.416

contact point points out of the VF, meaning that the VF tends to "pull away" from the contact point while opening. For this contact model to be valid, both the asymmetry in the VF vibration and the VF movement in the inferior-superior direction during closure need to be small. The former condition is confirmed in the simulations presented in Sec. III. The latter condition is valid as seen from the high-speed imaging of the VF vibration of the excised human larynx.[23]

The longitudinal stretching and 3D shear in the VFs during deformation due to anterior/posterior attachment to cartilage cannot be directly incorporated into the current 2D model. The authors therefore choose a set of elastic constants so that the lowest eigenmodes of the VFs resemble the corresponding 3D modes in the midplane presented by Luo *et al.*[19] where the physiological parameter ranges were adopted. In addition, the elastic constants are chosen so that when the subglottal pressure P_0 is in the physiological range, around 1 kPa, the glottal gap width would be in the physiological range too, i.e., on order of 1 mm. The viscoelastic model of the VFs is described by Luo *et al.*,[19] and here the authors only give a summary of the material properties. Each of the three VF layers is isotropic, and the shear modulus is $\mu=10$, 15, 5 kPa for the body, ligament, and cover, respectively. Poisson's ratio, density, and viscosity for all three layers are $\nu=0.3$, $\rho_s=1.0\ \text{g}/\text{cm}^3$, and $\eta=10$ P. The lowest four eigenmodes for the chosen parameters are identical to those of Luo *et al.*,[19] and the associated eigenfrequencies are 64, 151, 170, and 297 Hz. The eigenmodes have distinct patterns. The first mode corresponds to inclination of the VFs in the streamwise (x) direction, the second mode to stretching/compression in the transverse (y) direction, and the third and fourth modes represent the higher-order deflection forms.

The viscous incompressible Navier–Stokes equations are discretized using a second-order, centered finite-difference scheme in space and a second-order Crank–Nicolson scheme in time. The boundary conditions for both the flow and linear viscoelasticity of the VFs are treated with an IB method, which is described in detail by Mittal *et al.*[24] and Luo *et al.*[19] A non-uniform grid of 288×256 points is used for domain discretization with the minimum grid intervals $\Delta x=0.0234$ cm and $\Delta y=0.0078$ cm, and the time step size is $\Delta t=0.005$ ms for all simulations. The system is deemed to have reached a stationary state if the variations in the VF vibration amplitude are within 1%. A refined simulation has been performed with the grid points around the VFs doubled and Δt reduced by half. The simulation shows that changes in all the characteristics listed in Table I are below 10% and therefore confirms that the results are effectively independent of the grid resolution and time step size.

III. RESULTS AND DISCUSSIONS

Three simulations were performed with the subglottal pressure at $P_0=0.8$, 1.0, and 1.2 kPa. These values are above the phonation threshold, which is found to be around 0.5 kPa for the present model,[25] and they are within the physiological range.[26] Different values are chosen here to ensure that the obtained results can be generalized to a range of subglottal pressures, from somewhat above the onset threshold to 2.4 times as high. The authors present the simulation results for the VF deformation and flow physics separately in Secs. III A–III D.

A. VF deformation

Figure 2(a) shows a series of snapshots of the VF deformation immediately after the subglottal pressure, $P_0=1.2$ kPa, is imposed at time $t=0$. Due to the finite slope of the subglottal surface of the VFs, the suddenly imposed pressure produces not only a pushing force in the streamwise direction on the VFs but also a compressive force in the transverse direction. As a result, the VFs are displaced in both the x and y directions, and the glottis starts to open due to the VF deformation. Meanwhile, the flow is accelerated through the glottis and a jet is formed, the dynamics of

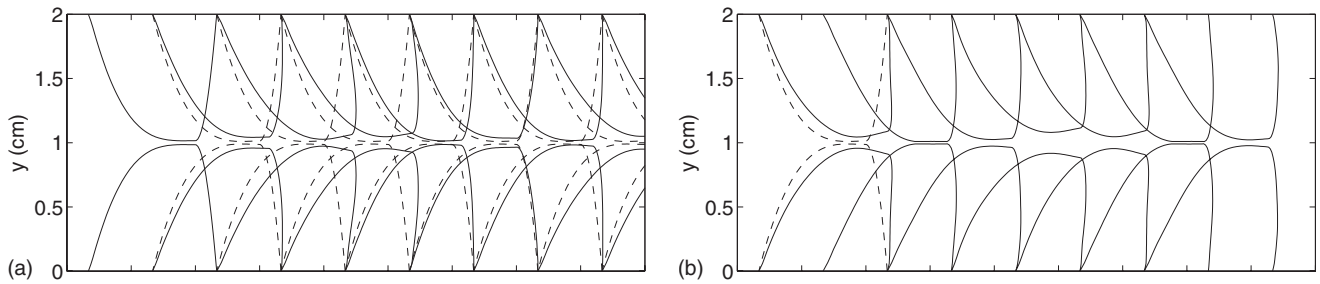


FIG. 2. A series of snapshots (shifted in the horizontal direction by 0.5 cm) of the deformed VFs, spaced by 1.5 ms, for $P_0=1.2$ kPa. (a) Transient vibration starting from $t=0$ and (b) sustained vibration. The undeformed VFs are shown as dashed lines.

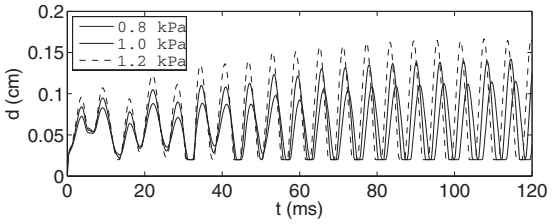


FIG. 3. Gap width between the VFs. $d(t)$.

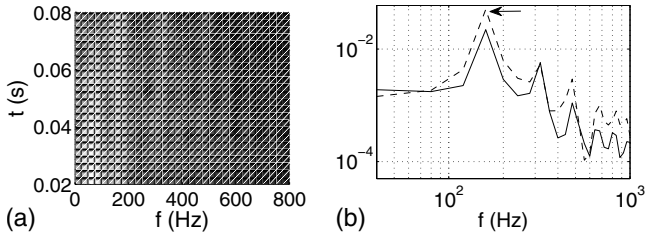


FIG. 4. (a) The time-frequency map during the transient stage. (b) The Fourier components of the x- (solid line) and y-coordinates (dashed line) of the VF superior tip during the sustained vibration.

which will be discussed in detail later. In the subsequent cycles, the VFs oscillate in both the x and y directions. Close inspection reveals that the streamwise and transverse motions are close to the first and second eigenmodes of the VF model, respectively. In Fig. 2(a), the frequency of the transverse deformation mode is approximately twice as high as the frequency of the streamwise deformation mode. This is indicated by the observation that the glottis opens and closes twice in one oscillation cycle. The VF deformation at this stage can be roughly described by the superposition of the two eigenmodes. The glottis formed by the two medial surfaces of the VFs changes shape, switching between convergent, straight, and divergent shapes. The convergent shape occurs when the closed VFs start to open from the subglottal side. From the eigenfunctions of the VFs,[19] neither the first nor the second eigenmode of the VFs contains such a deformation pattern. Therefore, the convergent shape in the present case could be caused by excitation of higher eigenmodes such as the third and fourth eigenmodes, which display the convergent shape of the glottis.

After a few cycles, the streamwise oscillation mode diminishes, and the VFs settle down to a displaced position in the x direction. Meanwhile, the transverse oscillation mode is amplified, and then this vibration mode is saturated and sustained for the cycles thereafter. Similar mode transition before the sustained vibration is also observed for $P_0=0.8$ and 1.0 kPa. Figure 3 shows the glottal gap width, d, defined as the minimal distance between the two VFs at any given instance, plotted against time for $P_0=0.8$, 1.0, and 1.2 kPa. In all cases, the glottal gap displays a transient stage that lasts for several cycles. The first few cycles contain double peaks, indicating that there are two oscillation frequencies present, which is consistent with the mode transition observed in Fig. 2(a). During the transition, the gap width gradually grows and finally settles down to a nearly periodic motion. Overall, the observed transient behavior can be roughly explained from a point of view of a general dynamical system. If the authors consider the coupled flow/VF an inhomogeneous system which is forced by the subglottal pressure and has a stable limit cycle, then the transient process can be viewed as a combination of the asymptotic decay of the initial condition and the asymptotic approach to the limit cycle. For the present system, both components of the process feature the eigenmodes of the VF structure. Finally, it should be pointed out that in this work, the authors only consider the situation where the VFs are initially at rest, and a constant subglottal pressure is suddenly imposed at $t=0$. In reality, the transient vibration can be complicated by the initial posture of the VFs and the pressure ramp in the trachea.

Figure 2(b) shows the self-sustained vibration pattern, demonstrated by a series of snapshots of the deformed VFs, for $P_0=1.2$ kPa. The maximum asymmetry in the deformation of the two VFs throughout the simulation is below 5% of the maximum glottal opening size. Therefore, the VF vibration is nearly symmetric. The glottis starts to open from the subglottal side as expected, and the two medial surfaces form a convergent shape blended in with the subglottal surfaces of the VFs. When the glottis is open, the majority portion of the medial surfaces forms either a straight or a divergent channel. Figure 2(b) shows that the glottis closes first from the subglottal side during the VF closing phase, which is consistent with the *in vivo* observation of the VF vibration[26] and the simulation of Thomson *et al.*[7] The geometric pattern of the glottis is an important topic in VF modeling, and the authors will discuss this issue in detail in Sec. III C. As the subglottal pressure is raised, the mean displacement of the VFs in the x direction becomes larger. As a result, the glottal opening size is widened, allowing a higher volume flow rate. In addition, the divergence angle between the two VFs becomes larger during the opening phase due to the increased inclination of the VF body. In the present simulations, the maximum included angles are around 18°, 24°, and 26° for $P_0=0.8$, 1.0, 1.2 kPa, respectively.

To better show the transient frequency components, the authors compute the time-frequency map of the x-displacement of the superior tip of the VFs using the short-time Fourier transform for the initial stage between $t=0$ and 100 ms. The map in Fig. 4(a), where $P_0=1.0$ kPa, displays two frequency bands around 70 and 163 Hz, which are close to the first two eigenfrequencies of the VF model. It can be seen that the 70 Hz component decreases in time, while the 163 Hz component increases in time. For sustained vibration at the same P_0, the frequency spectra are shown in Fig. 4(b), where the highest peak in the figure (indicated by an arrow) corresponds to the fundamental frequency at $f=163$ Hz, and the smaller peaks correspond to the harmonics of the fundamental frequency. The fundamental frequencies at the two other subglottal pressure levels are given in Table I and are discussed in Sec. III B. In all cases, the frequency is around 160 Hz and is slightly higher than the second eigenfrequency. During the stationary state (i.e., state of periodic vibration), the amplitude of the glottal gap width is $d=1.17$ mm for $P_0=0.8$ kPa, 1.42 mm for $P_0=1.0$ kPa, and 1.65 mm for $P_0=1.2$ kPa. These values are in the physiological range of human phonation where the amplitude is found to be on order of 1 mm.[27]

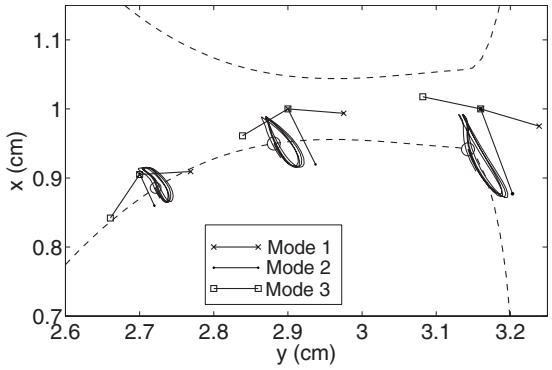


FIG. 5. The trajectories of three selected points on the VF during the stationary-state vibration for $P_0=1.2$ kPa. The start point is marked by a circle. The straight lines indicate the trajectories predicted by the first three eigenmodes (shifted from the VF for clarity).

Figure 5 shows the trajectories of the three selected points on the VF surface (as seen in Fig. 1) during the stationary-state vibration for $P_0=1.2$ kPa. A similar approach of analysis has been adopted previously.[28,29] All three trajectories show a nearly elliptical shape, and the points on the trajectories travel in the counterclockwise direction. These trajectories are in qualitative agreement with the recent observation of the vibration pattern of an excised human larynx.[23] Next to the ellipses are plotted the trajectories predicted by the first three eigenmodes of the VF model, which are simply straight lines. Note that the absolute lengths of the straight lines are arbitrary since the amplitude of the eigenmodes is undetermined. It can be seen that the elliptical trajectories are in general oriented along the lines corresponding to the second eigenmode, indicating that the flow-induced VF vibration pattern in the present case closely resembles this particular mode. The elliptical paths indicate that the deviation from the second eigenmode is mainly in the streamwise direction. This is caused by the streamwise oscillating load on the VFs during the VF opening/closure. Furthermore, the trajectories during the FSI are not strictly periodic. The aperiodic motion is closely related to the unsteady vortex motion in the flow, which will be discussed in Sec. III B.

The typical pressure force on one of the VF surfaces during a stationary-state vibration cycle is shown in Fig. 6(a) for $P_0=1$ kPa. The shear stress has been neglected when calculating the aerodynamic load on the VFs. This is because the dimensional analysis indicates that the shear stress is about one to two orders of magnitude lower than the pressure force during the VF vibration. The pressure fluctuation (pressure subtracted by its local temporal mean) is also shown here. The force on the medial surface of the VF is found to vary significantly during the cycle. When the glottis opens [first two frames in Fig. 6(a)], the pressure within the glottis reduces rapidly to zero (i.e., the atmospheric pressure level) along the streamwise direction. As the glottis opens further, the force on the medial surface diminishes. During this stage, the pressure produces a nominally compressive load on the VFs, forcing the glottis to open. Detailed analysis of the role of pressure will be provided in Sec. III C. When the glottis closes [right frame in Fig. 6(a)], the intraglottal pressure becomes negative, causing a local "pulling" force on the VF. On the subglottal surface of the VF, the pressure changes slightly around 1 kPa, while on the supraglottal surface, the temporal variation is notable.

B. Glottal jet

Figure 7 shows the volume flux per unit spanwise length, $Q=\int v_1 dy$, streamwise velocity v_1, and pressure p measured at the location $x=3$ cm on the centerline of the channel during the sustained vibration for the three cases. Note that the point of measurement is located within the glottis, as can be seen in Fig. 5. The volume flux has a pulsatile waveform whose amplitude increases as the subglottal pressure is raised. The statistical quantities of the volume flux are shown in Table I. These quantities include the maximum and mean volume flow rate, $Q_{\max}$ and Q_{mean}, fundamental vibration frequency $f=1/T$, where T is the period of 1 cycle, duration of the open phase T_o, rising time T_p and dropping time T_n of Q during the open phase, open quotient, $\tau_0=T_o/T$, skewness quotient, $\tau_s=T_p/T_n$, and the mean-to-maximum ratio of the volume flux, $q_r=Q_{\text{mean}}/Q_{\max}$. These quantities are either within the physiological range established in literature[26] or consistent with previous numerical studies of phonation.[13]

As P_0 is raised from 0.8 to 1.2 kPa, the volume flux increases by 74%, from 180 to 314 cm^2/s, and $Q_{\max}$ increases by 86%, from 405 to 755 cm^2/s. However, f changes only slightly from 162 to 164 Hz. Therefore, the subglottal pressure within the range considered here has little effect on the fundamental frequency of the VF vibration, and the fundamental frequency primarily depends on the inherent frequency of the VF structure, in the present case, the second eigenmode. The present relative independence of f on P_0 at the moderate frequency and subglottal pressure is consistent with the observation of Zemlin[26] and the simulations of Rosa *et al.*[12] Recent experiments using excised canine larynges

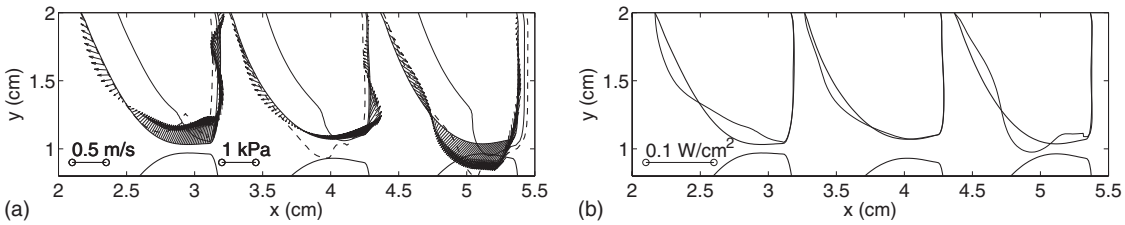


FIG. 6. (a) Velocity of the VF, pressure (thin solid line), and pressure fluctuation (multiplied by 2, dashed line) during a cycle for $P_0=1.0$ kPa where the three frames are spaced by 1.5 ms. (b) Corresponding energy flow intensity. The pressure indentation into the VF means positive p, while the energy indentation means positive energy flow into the VF.

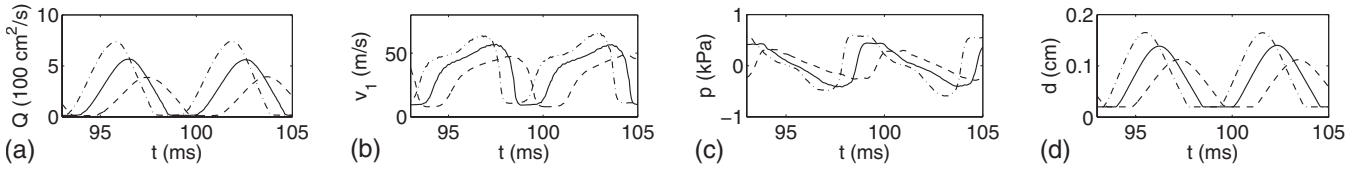


FIG. 7. (a) The volume flow rate Q, (b) streamwise velocity v_1, and (c) pressure p at $x=3.0$ cm and $y=1$ cm, and (d) glottal gap width d during sustained vibration for $P_0=0.8$ (dashed line), 1.0 (solid line), and 1.2 kPa (dashed-dotted line).

suggest that f has a nonlinear dependence on P_0.[30] The discrepancy between the simulations here and the experiments is probably caused by the linear VF model assumed in the present work. Due to the nonlinear nature of the dynamics, the fundamental frequency is shifted from the eigenfrequency by 8%. The $P_0=0.8$ kPa case has the largest open quotient among the three, meaning that the glottis has the longest open time. The $P_0=1.2$ kPa case has the highest skewness quotient but the lowest mean-to-maximum volume flux ratio.

The fluid velocity at the centerline shown in Fig. 7(b) has a quick rise as the glottis opens and then increases at a slower rate. It reaches a maximum during the closing phase and then reduces rapidly to a minimum. As P_0 is raised, the fluid velocity during the open phase increases. In addition, P_0 has a significant effect on the waveform of the fluid velocity. For $P_0=1.2$ kPa, there is a second quick rise in the velocity before the flow is shut off. The intraglottal pressure shown in Fig. 7(c) drops continuously during most of the open phase, and its minimum is well below zero. As P_0 is raised, the amplitude of the pressure oscillation increases, and the pressure waveform is also somewhat modified. The velocity and pressure waveforms presented here are qualitatively in agreement with the theoretical estimates by Titze.[31] In addition, the waveform of the fluid velocity is consistent with the experimental measurements of a mechanically driven model by Krane *et al.*[32] Finally, the general trends of the glottal pressure and velocity agree with the waveforms measured in excised canine larynges [e.g., Figs. 8(b) and 8(c) in Ref. 33]. Quantitative differences, however, do exist and are most likely due to the particular choices of the present VF geometry and material properties.

Throughout the VF vibration, the jet exhibits time-varying asymmetry and may attach itself to either one of the FVFs (Fig. 8). A similar phenomenon was also observed in previous experiments and numerical simulations.[3,10,15,19] As noted by Luo *et al.*,[19] the steady channel flow with a sudden expansion is known to bifurcate in its solution at a critical Reynolds number that depends on the expansion ratio.[18] Beyond the bifurcation point, the symmetric solution becomes unstable, and the steady flow becomes asymmetric even though the geometry is symmetric. In the present simulations, the peak Reynolds numbers based on the flow rate are $\mathrm{Re}_Q=(3/2)Q_{\max}/\nu_a=304$, 409, and 566 for $P_0=0.8$, 1.0, and 1.2 kPa, respectively, and the corresponding Reynolds numbers based on the centerline velocity and channel width are $\mathrm{Re}_c=4813$, 5625, and 6375. These Reynolds numbers are much higher than the critical Reynolds number at the present expansion ratios.[18] Therefore, the observed flow asymmetry is a natural effect of the present larynx model. Further discussion on this issue is provided in Sec. III D.

The downstream flow during the stationary-state vibration is dominated by asymmetric and unsteady vortex motions, as seen in Fig. 8 where two snapshots of typical instantaneous velocity field downstream the FVFs are shown for $P_0=1.2$ kPa. For clarity, the velocity vectors from every four grid points in the x direction and every five points in the y direction are shown. In Fig. 8(a), there are a pair of large counter-rotating vortices downstream the FVFs that are staggered in the channel. The jet downstream the VFs is clearly deflected by the two vortices. Between the FVFs, smaller vortices can be seen, and the jet is directed toward one of the FVFs. The large vortices in the channel evolve in time, changing their size and center locations, and some of them are strengthened by the jet. Consequently, the downstream portion of the jet changes its shape and appears to be flapping. Away from the glottis, the jet core is widened due to the momentum exchange between the jet core and the ambient fluid. As the VFs close, the mean flow diminishes, but the flow in the supraglottal region continues to circulate due to the inertial effect. During this period, the smaller vortices in the flow may disappear due to the viscous dissipation or merge into larger vortices. The surviving vortices will affect the jet behavior in the next cycle. Figure 8(b) shows the velocity field from another cycle where the jet is diverted

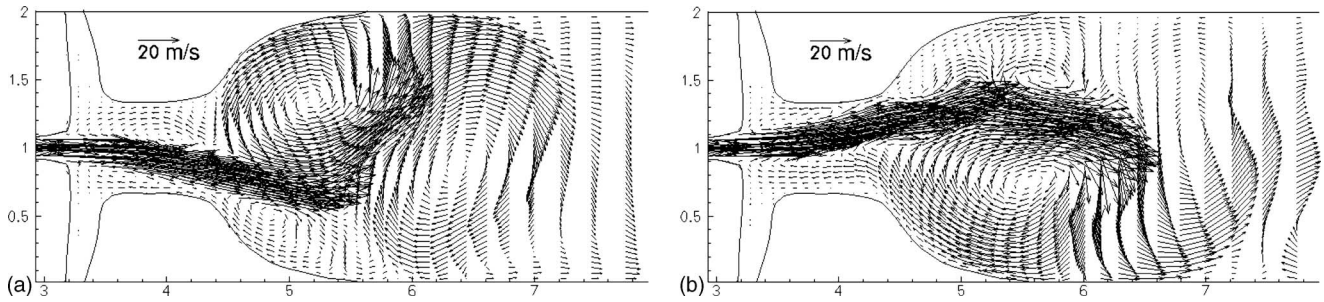


FIG. 8. Two typical instantaneous flow fields for $P_0=1.2$ kPa.

toward the other FVF by a clockwise vortex and attaches to that FVF. In all the cycles, the skewness of the jet in the region between the FVFs is relatively persistent for the cycle, while downstream the FVFs, the jet skewness changes rapidly due to the jet-vortex interaction.

Flow separation is of interest during phonation as it significantly affects the pressure distribution on the VF surfaces.[3,34] Since the glottis may form a divergent channel in the open phase, the flow between the two VFs may separate from the medial surfaces. This phenomenon was studied in the numerical simulations of the stationary VF model[10] and in the experimental study of the glottal jet.[17] In the present simulations, flow separation from the medial surfaces is observed for P_0=1.0 and 1.2 kPa. For P_0=0.8 kPa, recirculation within the glottis is not obvious. The separation zone usually appears between the jet and the medial surface of one VF when the jet is deflected toward the other VF. Throughout the simulation, the location of the separation point and size of the separation zone vary according to the shape of the glottis and deflection of the jet. For example, when the VFs close, the separation point moves from the medial surface to the glottal exit, and in next cycle, it may re-appear on the same side of the glottis or switch to the other side. The similar time-varying separation was also reported in the experimental observation by Neubauer *et al.*[17]

C. Pressure work and energy

Some laryngoscopic studies[26] suggest that the glottis may change from a convergent shape during the opening phase to a divergent shape during the closing phase. This vibratory feature has been adopted into many previous VF modeling works, especially those employing two-mass models,[3] which necessarily generate a phase difference between the aerodynamically driven mass and the passively following mass. In Ref. 7, the authors argued that a convergent-divergent pattern is necessary for a net amount of energy to be transferred from the fluid to the VFs during a whole vibration cycle and thus for the VFs to maintain the passive vibration. According to their argument, the convergent glottis during the VF opening maintains an intraglottal pressure which is higher than the pressure at the glottal exit and does positive work to the VFs, and the divergent glottis during the VF closing maintains a lower intraglottal pressure which does less amount of work to the VFs than during the opening phase. This temporal asymmetry in the pressure ensures that the net energy flow to the VFs in one full cycle is positive.

In the present simulations, the stationary-state VF vibration pattern also contains a convergent-divergent pattern. Note that, as shown in Fig. 2(b), a portion of the glottis forms a convergent shape during the opening phase that is blended in with the subglottal surface. However, as pointed out by Titze,[31] the air inertia may be another factor to produce the similar pressure asymmetry to sustain the energy flow. In agreement with his theory, Fig. 7 shows that the fluid velocity in the glottis is not completely in phase with the glottal gap width. When the glottis opens at the beginning of a cycle, the velocity of the subglottal fluid lags behind the glottal gap due to the inertia of the subglottal fluid as well as the resistance from the supraglottal fluid. Similarly, during the closing phase, the fluid velocity does not drop concurrently with the glottal gap since the already-established jet tends to maintain its momentum. In fact, Fig. 7(b) shows that the intraglottal velocity reaches its maximum when the VFs close. This assertion on the role of the fluid inertia is in line with the argument by Krane *et al.*[32] in explaining their experimental measurement of the glottal velocity at the centerline, which has a similar waveform as shown in Fig. 7(b). Corresponding to the velocity waveform, the intraglottal pressure is not in phase with the glottal gap width but is high during the opening phase and low during the closing phase, as suggested in Figs. 6(a) and 7(c). In conclusion, a convergent-divergent vibration mode may not be necessary to maintain the net energy transfer to the VFs or the sustained vibration.

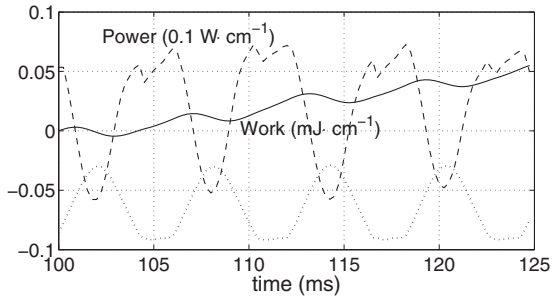


FIG. 9. Work (solid line) and the rate of work (dashed line) done to one of the VFs by the flow pressure during the sustained vibration for P_0 =1.0 kPa. The glottal gap width d is represented as the thick-dotted line (shifted in ordinate for clarity).

To show the energy transfer between the flow and VFs, the authors compute the total power and work done on one of the VFs by the flow pressure during the sustained vibration. The work done by the shear stress on the VF surface has been neglected since it is very small compared to the work done by the pressure.[7] Figure 9 shows that the difference of the work between any two points separated by a full cycle period is always positive, indicating that positive net energy is transferred from the flow to the VF. Note that alternatively one could analyze the work done by the pressure fluctuation [shown in Fig. 6(a)] only, as the net work done by the temporal mean pressure is zero. The rate of the work varies its sign and is not in phase with the glottal gap width d. During the VF closure and early stage of the following opening phase, the power to the VF is positive, meaning that the VFs are forced to deform. The typical velocity of the VFs and the corresponding pressure distribution at this stage are shown in the left frame of Fig. 6(a), and the corresponding energy flow intensity (the dot product of the pressure force and VF velocity) is shown in the left frame of Fig. 6(b). The power is computed by integrating the energy flow intensity over the entire VF surface. The power later becomes negative for a time period that involves both the opening phase and closing phase (Fig. 9). During the time period, the pressure force works against the VFs, either preventing them from further deformation or delaying their elastic recoil. It is interesting that the power in Fig. 9 reaches a negative minimum when

the glottis has nearly maximum opening. To explain this, the authors plot the velocity of the VF and the pressure load when the glottal gap width reaches its maximum, as shown in the middle frame of Fig. 6(a). It can be seen that the VF continues to deform even when d reaches its maximum value, and the subglottal surface of the VF moves against the high pressure, which results in a negative energy flow into the VF [middle frame of Fig. 6(b)].

During the late stage of the closing phase, the power to the VF becomes positive again (Fig. 9). This indicates that the pressure force assists the VFs to close. The typical VF velocity, pressure force, and energy flow at this stage are shown in the right frames of Figs. 6(a) and 6(b). Note that beside the medial surface, where the energy flow intensity is positive, there is a significant part of the subglottal surface where the energy flow is positive. That is, the VFs are being pulled at the medial surface while being pushed at the subglottal surface. Therefore, the VFs close not only under the interior elastic recoil but also with the assistance of the pressure. This role of the pressure was not observed by Thomson *et al.*[7] In their simulations, Thomson *et al.* found that overall, the pressure force never assists the VFs during closure even though there is a negative pressure locally at the medial surface pulling the VFs. To reconcile the difference between the results here and Thomson *et al.*, the present authors point out that the geometric model of the VFs here is significantly different from that of Thomson *et al.*, and as a result, the subglottal surface of their VF model has less movement compared to the present model during the vibration. In conclusion, the exact role of pressure in the VF vibration depends on the chosen model, and it is possible that the pressure force works cooperatively with the elastic force in the VFs to close the glottis.

D. Mechanism for the flow asymmetry

In the present study, both the geometry of the larynx model and VF material properties are symmetric, but the glottal jet and downstream vortices exhibit highly asymmetric behavior. The attachment of the flow to one of the VF medial surfaces and deflection of the jet core have been observed in both experiments and numerical simulations.[3,10,15] However, this asymmetry was explained by some of the authors as resulting from the so called Coanda effect.[3,15] According to the argument, when the flow goes through the glottis with a divergent cross section, the fluid trapped between the main stream and one of the VF surfaces is entrained to move, and the evacuating effect combined with the high pressure on the other side of the flow would deflect the flow and attach it to the VF. However, the present simulations show that the jet flow can be severely deflected even when the flow is symmetric within the glottis, as seen in Fig. 8. In addition, the largest curvature of the deflected jet is not in the glottis but somewhere downstream in the supraglottal region. Therefore, the result here is more consistent with the study of the steady symmetric inflow entering a sudden expansion by Drikakis,[18] where the flow with the expansion ratio of 10 becomes asymmetric at a very low critical Reynolds number $\mathrm{Re}_Q=26$. According to Drikakis,[18] the critical Reynolds number decreases as the expansion ratio is increased. His result suggests that the downstream flow confined by the channel plays a key role in the flow asymmetry. Details of the flow structures presented in the present work reveal that the jet has a strong interaction with the vortices generated due to the confinement of the channel walls. During the interaction, the upstream portion of the jet is deflected by the vortices toward either wall of the FVFs, and the downstream portion displays a wavy form. On the other hand, the jet may strengthen the vortices and cause them to roll up. As a result, the skewness of the jet is time-varying within one vibration cycle, and especially further downstream, the curvature of the jet varies drastically due to the rapid evolution of the unsteady vortices. Note that the similar jet/vortex interaction is expected to occur in a corresponding 3D flow, even though the details may be different from those presented in Sec. III B due to the 3D vortex structures. Durst *et al.*[35] showed that a 3D flow over a plane symmetric sudden expansion also becomes asymmetric above a critical Reynolds number. Given that the jet Reynolds number of the glottal flow varies during a vibration cycle, the extent of flow asymmetry should also change, leading the 3D jet to oscillate.

The wavy pattern and flapping motion of the jet have been reported by Neubauer *et al.*[17] in their experimental visualization of the flow downstream a self-oscillating VF model where the Reynolds number and channel expansion ratio are much higher than those in the present work. Therefore, the present findings are consistent to their results. Neubauer *et al.*[17] hypothesized that the jet asymmetry is induced by the flow structures downstream from the VFs. This hypothesis is confirmed by the numerical simulations in this work.

IV. CONCLUSIONS

A recently developed IB method implemented on fixed Cartesian grids is used to solve the governing equations for both the incompressible flow and linear viscoelasticity of the VF model. Compared to the previous research on phonation modeling, the present work advances the use of continuum-based model by using an accurate and efficient method and captures features of the VF vibration and glottal flow with a high level of detail. The transient vibration during onset, the role of pressure in energy transfer, and the oscillatory behavior of the glottal jet have been focused on in this study.

The simulations were performed with three different subglottal pressure levels above the phonation threshold. Self-sustained vibrations of the VFs initially at rest are obtained, and the temporal patterns of the glottal opening and measures of the intraglottal velocity and pressure are reasonably realistic. The results show that the VF vibration undergoes a distinct transition at the start of phonation. The transient stage is dominated by the mode that closely resembles the first mode identified in the eigenmode analysis, whereas the stationary state is identified more closely with the second eigenmode. As expected, the net work done by the flow pressure to the VFs is positive in one vibration cycle. However,

the pressure plays distinct roles during different phases of a cycle, and it may promote the VF closure during the closing phase.

The flow structures reveal that in the downstream region away from the VFs, the flow is dominated by unsteady large-scale vortices. The glottal jet interacts with these vortices, and consequently, the direction of the jet is deflected and displays dramatic cycle-to-cycle oscillation. In addition, the downstream part of the jet in the present model appears as a wavy form due to the counter-rotating vortex pairs, and its shape is rapidly altered by the vortices.

Although the present work is limited to a 2D study and the 3D effect is not incorporated, it sheds light on the mechanism of selecting a particular vibration mode from several structural eigenmodes by the nonlinear FSI process and also on the complex flow behavior. Work is currently underway to investigate the corresponding issues in 3D models.

ACKNOWLEDGMENTS

This work was supported by Grant No. R0lDC007125 from the NIDCD. We would like to thank Dr. Anders Löfqvist and the two anonymous reviewers for many helpful comments.

[1]Y. Zhang, C. McGilligan, L. Zhou, M. Vig, and J. J. Jiang, "Nonlinear dynamic analysis of voices before and after surgical excision of vocal polyps," J. Acoust. Soc. Am. **115**, 2270–2277 (2004).
[2]J. J. Jiang, Y. Zhang, and J. Stern, "Modeling of chaotic vibrations in symmetric vocal folds," J. Acoust. Soc. Am. **110**, 2120–2128 (2001).
[3]C. Tao, Y. Zhang, D. G. Hottinger, and J. J. Jiang, "Asymmetric airflow and vibration induced by the coanda effect in a symmetric model of the vocal folds," J. Acoust. Soc. Am. **122**, 2270–2278 (2007).
[4]D. A. Berry and I. R. Titze, "Normal modes in a continuum model of vocal fold tissues," J. Acoust. Soc. Am. **100**, 3345–3354 (1996).
[5]D. A. Berry, H. Herzel, I. R. Titze, and K. Krischer, "Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions," J. Acoust. Soc. Am. **95**, 3595–3604 (1994).
[6]F. Alipour, D. A. Berry, and I. R. Titze, "A finite-element model of vocal-fold vibration," J. Acoust. Soc. Am. **14**, 442–454 (2000).
[7]S. L. Thomson, L. Mongeau, and S. H. Frankel, "Aerodynamic transfer of energy to the vocal folds," J. Acoust. Soc. Am. **118**, 1689–1700 (2005).
[8]C. Tao and J. J. Jiang, "Anterior-posterior biphonation in a finite element model of vocal fold vibration," J. Acoust. Soc. Am. **120**, 1570–1577 (2006).
[9]F. Alipour and R. C. Scherer, "Flow separation in a computational oscillating vocal fold model," J. Acoust. Soc. Am. **116**, 1710–1719 (2004).
[10]R. C. Scherer, D. Shinwari, K. J. De Witt, C. Zhang, B. R. Kucinschi, and A. A. Afjeh, "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees," J. Acoust. Soc. Am. **109**, 1616–1630 (2001).
[11]B. R. Kucinschi, R. C. Scherer, K. J. DeWitt, and T. T. M. Ng, "An experimental analysis of the pressures and flows within a driven mechanical model of phonation," J. Acoust. Soc. Am. **119**, 3011–3021 (2006).
[12]M. O. Rosa, J. C. Pereira, M. Grellet, and A. Alwan, "A contribution to simulating a three-dimensional larynx model using the fine element method," J. Acoust. Soc. Am. **114**, 2893–2905 (2003).
[13]C. Duncan, G. Zhai, and R. C. Scherer, "Modeling coupled aerodynamics and vocal fold dynamics using immersed boundary methods," J. Acoust. Soc. Am. **120**, 2859–2871 (2006).
[14]R. S. McGowan, "An aeroacoustic approach to phonation," J. Acoust. Soc. Am. **83**, 696–704 (1988).
[15]B. D. Erath and M. W. Plesniak, "An investigation of bimodal jet trajectory in flow through scaled models of the human vocal tract," Exp. Fluids **40**, 683–696 (2006).
[16]M. Triep, C. Brüker, and W. Schröder, "High-speed PIV measurements of the flow downstream of a dynamics mechanical model of the human vocal folds," Exp. Fluids **39**, 232–245 (2005).
[17]J. Neubauer, Z. Zhang, R. Miraghaie, and D. A. Berry, "Coherent structures of the near field flow in a self-oscillating physical model of the vocal folds," J. Acoust. Soc. Am. **121**, 1102–1118 (2007).
[18]D. Drikakis, "Bifurcation phenomena in incompressible sudden expansion flows," Phys. Fluids **9**, 76–87 (1997).
[19]H. Luo, R. Mittal, X. Zheng, S. Bielamowicz, R. Walsh, and J. Hahn, "An immersed-boundary method for flow-structure interaction in biological systems with application to phonation," J. Comput. Phys. **227**, 9303–9332 (2008).
[20]A. Cooke, C. L. Ludlow, N. Hallet, and W. S. Selbie, "Characteristics of vocal fold adduction related to voice onset," J. Voice **11**, 12–22 (1997).
[21]Z. Zhang, J. Neubauer, and D. A. Berry, "Physical mechanisms of phonation onset: A linear stability analysis of an aeroelastic continuum model of phonation," J. Acoust. Soc. Am. **122**, 2279–2295 (2007).
[22]Z. Zhang, "Influence of flow separation location on phonation onset," J. Acoust. Soc. Am. **124**, 1689–1694 (2008).
[23]A. Boessenecker, D. A. Berry, J. Lohscheller, U. Eysholdt, and M. Doellinger, "Mucosal wave properties of a human vocal fold," Acta Acust. **93**, 815–823 (2007).
[24]R. Mittal, H. Dong, M. Bozkurttas, F. M. Najjar, A. Vargas, and A. von-Loebbeck, "A versatile sharp interface immersed boundary method for incompressible flows with complex boundaries," J. Comput. Phys. **227**, 4825–4852 (2008).
[25]X. Zheng, "Biomechanical modeling of glottal aerodynamics and vocal fold vibration during phonation," Ph.D. thesis, George Washington University, Washington, DC (2009).
[26]W. R. Zemlin, *Speech and Hearing Science: Anatomy and Physiology*, 3rd ed. (Prentice-Hall, Englewood Cliffs, NJ, 1988).
[27]S. Schuberth, U. Hoppe, M. Döllinger, U. Eysholdt, and J. Lohscheller, "High-precision measurement of the vocal fold length and vibratory amplitudes," Laryngoscope **112**, 1043–1049 (2002).
[28]T. Baer, "Observations on vocal fold vibration: Measurement of excised larynges," in *Vocal Fold Physiology*, edited by K. N. Stevens and M. Hirano (University of Tokyo Press, Tokyo, 1981), pp. 119–133.
[29]H. Fukuda, Y. Kawasaki, M. Kawaida, A. Shiotani, K. Oki, T. Tsuzuki, T. Fujioka, and E. Takayama, "Physiological properties and wave motion of the vocal fold membrane viewed from different directions," in *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Mechanisms*, edited by J. Gauffin and B. Hammarberg (Singular, San Diego, CA, 1991), pp. 7–14.
[30]F. Alipour and R. C. Scherer, "On pressure-frequency relations in the excised larynx," J. Acoust. Soc. Am. **122**, 2296–2305 (2007).
[31]I. R. Titze, "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**, 1536–1552 (1988).
[32]M. Krane, M. Barry, and T. Wei, "Unsteady behavior of flow in a scaled-up vocal folds model," J. Acoust. Soc. Am. **122**, 3659–3670 (2007).
[33]F. Alipour and R. C. Scherer, "Pulsatile airflow during phonation: An excised larynx model," J. Acoust. Soc. Am. **97**, 1241–1248 (1994).
[34]D. Sciamarella and P. Le Quéré, "Solving for unsteady airflow in a glottal model with immersed moving boundaries," Eur. J. Mech. B/Fluids **27**, 42–53 (2008).
[35]F. Durst, A. Melling, and J. H. Whitelaw, "Low Reynolds number flow over a plane symmetric sudden expansion," J. Fluid Mech. **64**, 111–128 (1974).

Vowel and consonant contributions to vocal tract shape

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

(Received 22 October 2008; revised 13 May 2009; accepted 13 May 2009)

The purpose of this study was to develop a method by which a vowel-consonant-vowel (VCV) utterance based on x-ray microbeam articulatory data could be separated into a vowel-to-vowel transition and a consonant superposition function. The result is a model that represents a vowel sequence as a time-dependent perturbation of the neutral vocal tract shape governed by coefficients of canonical deformation patterns. Consonants were modeled as superposition functions that can force specific portions of the vocal tract shape to be constricted or expanded, over a specific time course. The three VCVs [əpɑ], [ətɑ], and [əkɑ], produced by one female speaker, were analyzed and reconstructed with the developed model. They were shown to be reasonable approximations of the original VCVs, as assessed qualitatively by visual inspection and quantitatively by calculating rms error and correlation coefficients. This establishes a method for future modeling of other speech material. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3158816]

PACS number(s): 43.70.Bk, 43.70.Jt, 43.70.Aj [CHS] Pages: 825–836

I. INTRODUCTION

It is well known that speech does not merely consist of a serial concatentation of individual speech sounds, but rather a continuous blending of one sound into another. Production of a blended acoustic stream requires that movements of the articulatory structures be temporally and spatially coordinated such that they are engaged in generating more than one sound segment at any instant of time. This process of overlapping speech movements, commonly referred to as *coarticulation* (e.g., Kozhevnikov and Chistovich, 1965; Öhman, 1966; Kent and Minifie, 1977) or *coproduction* (e.g., Fowler, 1980), must be included in the development of any realistic model of speech production. Because movement of the articulators has the collective effect of determining the *shape* of the vocal tract airspace, and because the shape of the airspace is the most direct connection to the acoustic resonances expressed in the speech signal as formants, it is important to develop a model that directly relates overlap of phonetically-relevant gestures to the time-dependent changes of the tract shape during production of speech. The purpose of this study was to determine whether realistic changes in vocal tract shape can be accurately represented with a model where coarticulated speech is produced by superimposing consonant gestures on an ongoing sequence of vowel transitions.

This particular view of coarticulation seems to have originated with Öhman's (1966) spectrographic analysis of vowel-consonant-vowel (VCV) utterances. The results suggested that vowels and consonants are generated by two parallel production systems and place coincident demands on the same articulatory structures. Thus, a VCV is considered to be produced as a vowel-vowel (VV) sequence upon which a consonant gesture is superimposed. Based on a cineradiographic study of /hə'CV/ utterances, Perkell (1969) similarly concluded that production of vowels and consonants is carried out by two functionally different types of articulatory activity. He noted that consonant articulations are generally faster and require more precise timing than vowel articulations. This led to the hypothesis that articulatory activity could be divided into two classes in which the extrinsic speech musculature is utilized primarily for vowel production, whereas the intrinsic musculature of the tongue and lips executes the place, manner, and degree of a specific consonant that is additively superimposed on the overall vocal tract posture provided by the extrinsic musculature. That is, rapid, precise consonantal gestures are superimposed on an underlying vowel gesture. Gracco (1992) also referred to a functional division of vocal tract movement into two general categories: "…those that produce and release constrictions (valving), and those that modulate the shape or geometry of the vocal tract." Löfqvist and Gracco (1999) later reported that, during production of a VCV, tongue movement toward the final vowel often began before the consonant closure had occurred, and further that most of the movement from the initial vowel occurred during the consonant closure. Both findings suggest that the overall shaping of the vocal tract continues while constrictions are imposed and released.

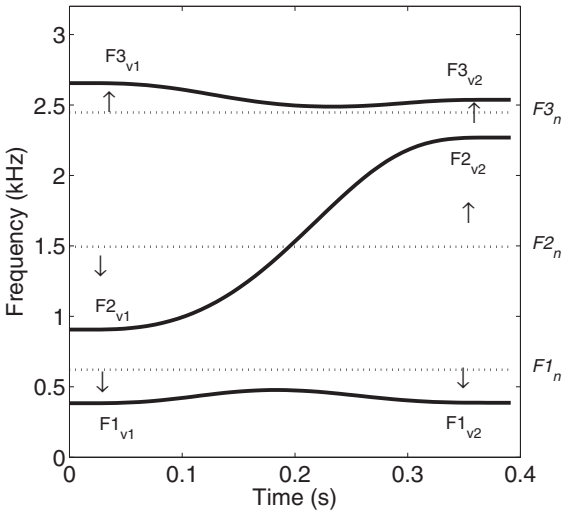
The paradigm of functional division of the articulatory system into separate vowel and consonant classes has influenced various control strategies for articulatory and vocal tract models. Öhman (1967) followed his spectrographic study by proposing a model that allows for interpolation of the midsagittal cross-distance (width) of one vowel shape to another over the time course of an utterance. Simultaneously, a consonant constriction function can be activated that varies over the same time course as the vowel component. At each successive point in time, the consonantal function is superimposed on the modeled vowel substrate to produce a composite tract shape. Nakata and Mitsuoka (1965) and Ichikawa and Nakata (1968) used the idea of superimposing a consonant on a VV transition in a rule-based speech synthesizer.

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

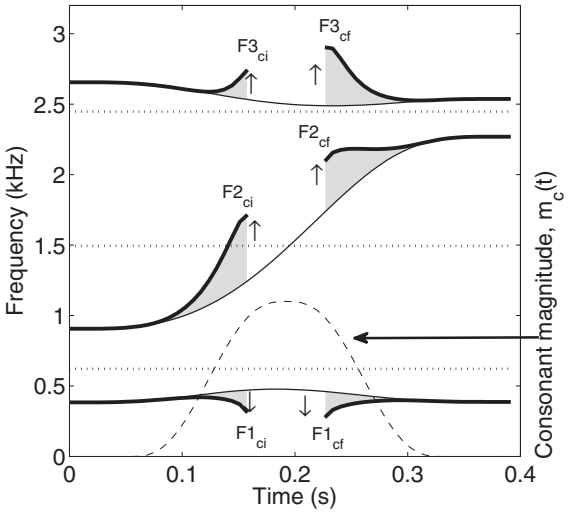
Similarly, Båvegård (1995) and Carré and Chennoukh (1995) both reported vocal tract area function models where consonant constrictions are superimposed on an interpolation of a vowel-to-vowel transition. Browman and Goldstein's (1990) development of "articulatory phonology" also reflects the ideas suggested by Öhman's work. In their view, speech is produced by a series of overlapping gestures created by activation of "tract variables" such as constriction location and degree of the tongue body and tip. Models based on similar conceptualizations of speech production have been proposed by Fowler and Saltzman (1993), Byrd (1996), and Tjaden (1999).

More recently, Story (2005a) described a model of the vocal tract area function that incorporates many aspects of the type of speech production system suggested by Öhman (1966, 1967) and Perkell (1969). The model consists of multiple hierarchical tiers,[1] each of which is capable of imposing particular types of perturbation on the vocal tract shape. In the first tier, vowels and vowel-to-vowel transitions are produced by perturbing a phonetically-neutral vocal tract configuration with two canonical deformation patterns called modes. The modes, which were derived by principal component analysis, are shaped such that they efficiently exploit the acoustic properties of the vocal tract (Story, 2007a), and tend to be common across speakers (Story and Titze, 1998; Story, 2005b, 2007b). Consonant production results from a second tier of perturbation that imposes severe constrictions on the underlying vowel or evolving VV transition. The constriction functions are precisely defined by the range along the tract length that is affected by the constriction, the constriction location, the cross-sectional area at the point of maximum constriction, and by the timing of the activation.

A speech signal produced by this model[2] carries information characteristic of the successive tiers of vocal tract structure or movement on which it is built. For example, the idiosyncratic features of the neutral vocal tract shape will set the acoustic background on which vowel transitions generated by the first tier are superimposed. This is demonstrated in Fig. 1(a) where time-dependent formant contours are shown for a transition from [ʊ] to [i] simulated by the model in terms of area function changes. At any point in time, the formants along the transition can be considered to be perturbed or deflected away (as suggested by the up and down arrows) from the formants of the underlying neutral tract (shown with dotted lines). The characteristics of the vowel transitions, in turn, provide the acoustic background on which consonantal perturbations are imposed by the second tier of the model. Demonstrated in Fig. 1(b) are the formant contour effects that result from perturbing the [ʊi] transition with a consonant constriction intended to approximate a [d]. For comparison purposes, the [ʊi] transition and neutral tract formants are also shown in this figure, along with the time course of the consonant magnitude, denoted as $m_c(t)$. The consonant magnitude indicates the period of time in which the consonant perturbation affects the vocal tract shape [i.e., for $m_c(t)>0$] and causes the formant frequencies to be deflected upward or downward relative to those formants that would exist in the absence of the consonant (see arrows).



(a) [ʊi]



(b) [ʊdi]

FIG. 1. Demonstration of multiple "layers" of vocal tract perturbations to produce a VCV. (a) Time-varying formant frequencies (thick solid lines) that result from perturbing a neutral vocal tract shape with vowel-to-vowel (VV) sequence; the F_n's are the formants for the neutral tract shape and shown with dotted lines, F_v's are the formants in an approximately stable portion of the initial and final vowels. (b) Time-varying formant frequencies (thick solid lines) that result from perturbing the VV sequence (thin solid lines) with a consonant perturbation approximately representative of [d]; the dashed line indicates the time course of the consonant activation whose amplitude ranges from 0 to 1 (not Hz) and the F_{ci}'s and F_{cf}'s are formant frequencies that exist at the point where the occlusion begins and ends.

Although utterances can be simulated with this model, temporal activations of the vowel and consonant elements have been mostly prescribed by *ad hoc* rules (Story, 2005a) or formant-to-coefficient mapping (Story and Titze, 1998). It was shown in Story (2007b), however, that canonical modes similar to those on which the vowel tier is based could be extracted from the articulatory data available in the University of Wisconsin-Madison's x-ray microbeam (XRMB) database (Westbury, 1994). Specifically, the *x*-*y* coordinates of fleshpoint pellets, along with the outline of the hard palate, were used to approximate the shape of the oral part of the vocal tract as a midsagittal *cross-distance* function for eleven

vowels. A principal component analysis (PCA) performed on speaker-specific sets of these cross-distance functions revealed mode shapes essentially the same as those derived from MRI-based (magnetic resonance imaging) area function sets in previous research (e.g., Story and Titze, 1998; Story, 2005b; Mokhtari *et al.* 2007), even though only the oral portion of the vocal tract was available for analysis. Further analysis demonstrated that VV transitions could be accurately represented with time-dependent scaling of the modes superimposed on the mean cross-distance function, thus validating the first tier of the Story (2005a) model. It remains a question though whether *naturally spoken* VCVs can be accurately represented by modeling them as a VV transition with a superimposed consonant function, as prescribed by the second tier of the model.

The aim of this study was to address this question by extracting time-varying cross-distance functions from XRMB data, with the method described in Story (2007b), for VCV utterances with stop consonants, and determining how well their time course can be represented with a model similar to Story (2005a) but formulated for cross-distance, rather than area. The assumption was that the initial and final vowels in a VCV can be well represented by scaling coefficients of the vowel-based modes. These coefficients can then be interpolated to generate the underlying VV transition. It was also assumed that the consonant superposition function can be determined from the shape of the vocal tract during the closure period, and represented with parameters such as constriction location, extent, degree, and temporal activation.

Described in Secs. II–V is a step-by-step process for transforming the *x*-*y* coordinates of articulatory fleshpoints during production of a VCV to the components of a vocal tract model based on superimposing a consonant gesture on a vowel substrate. Since the outcome of early steps in the process are needed for later steps, each section combines explanations of the method with intermediate results.

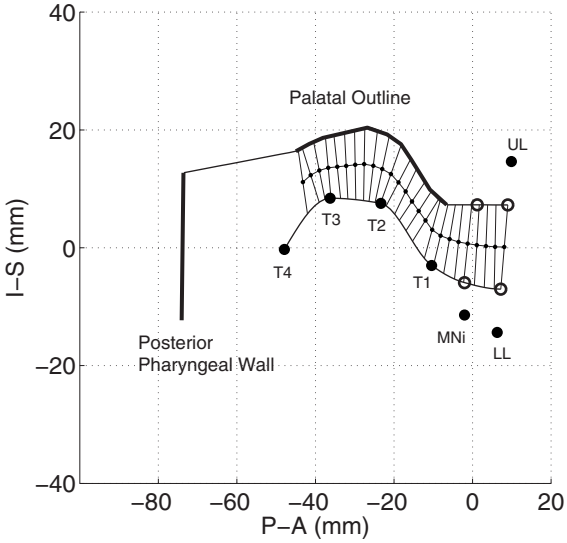
II. MEASUREMENT OF TIME-VARYING CROSS-DISTANCE FUNCTIONS

A. Speaker and speaking tasks

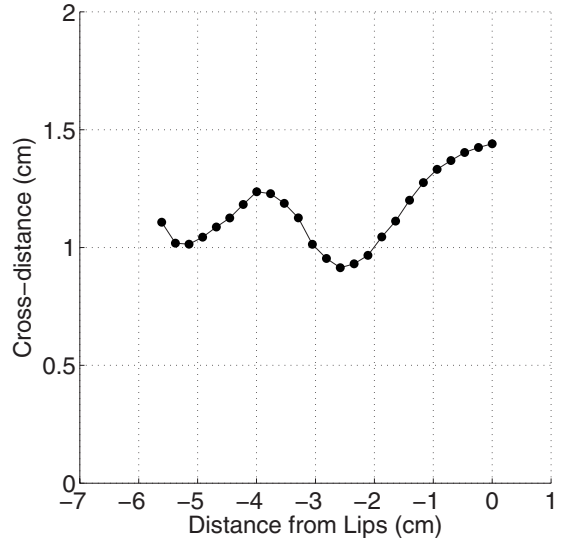
The speaker chosen from the XRMB database for the present study was JW26. This female speaker was 24 years old at the time of data collection and her dialect base is listed as Verona, Wisconsin (Westbury, 1994). She was also one of the four speakers studied in Story (2007b) so that vocal tract modes have already been reported for her vowels. These will be used subsequently in Sec. III. The XRMB protocol for each speaker in the database contains a set of VCVs in the form of [əCɑ], where "C" is 1 of 21 different consonants (task number 16, file "tp016"). Only those VCVs with the unvoiced stop consonants [p, t, k] were analyzed in this study.

B. Cross-distance functions from XRMB data

The XRMB data consist of time-dependent displacements of gold pellets affixed to four points on the tongue, two on the jaw, and one on each of the upper and lower lips. During data collection, *x*-*y* coordinates of each pellet were acquired at a sampling interval of 6.866 ms (Westbury, 1994). To extract a representation of the vocal tract shape, the same technique was used as described by Story (2007b) where a midsagittal cross-distance function can be determined from a two-dimensional vocal tract profile, constructed from pellet coordinates, for any time frame of data produced by a given speaker. Although the details of the technique will not be repeated here, the process is demonstrated graphically in Fig. 2. The positions of the pellets on the tongue (T1–T4), incisor (MNi), lower lip (LL), and upper lip (UL) are shown as filled circles in Fig. 2(a) for a specific time frame representative of JW26's [æ] vowel. Also shown



(a)



(b)

FIG. 2. Demonstration of finding a cross-distance function from XRMB data. (a) Sagittal view of a time frame representative of JW26's [æ] vowel. A superior and an inferior vocal tract boundary are generated based on the tongue points (T1-T4), palatal outline, pharyngeal wall, and four phantom points (open circles) related to the mandible and lips. The lines extending across the vocal tract are perpendicular to the centerline and comprise the cross-distance measurements. (b) Resulting cross-distance function; note that the *x*-axis indicates distance into the vocal tract from the lips, where the lips are at 0 cm.

is the palatal outline and an approximation of the posterior pharyngeal wall. The open circles denote derived "phantom" points (see Story, 2007b, p. 3774) that more accurately represent the interface of tissue and air than do the actual pellets on the lips and jaw.[3] The inferior vocal tract profile is generated with a spline fit through the four tongue points and the two lower phantom points. Similarly, the upper profile is constructed with a spline fit through the upper two phantom points and the palatal outline. The next step is to use an iterative bisection algorithm to find the centerline through the airspace, extending from the lips back to the most posterior point of the palatal outline. Finally, the distance from the lower to upper profile is measured perpendicularly at a succession of points along the centerline. This collection of cross-distances can be plotted as a function of the distance from the lips, as shown in Fig. 2(b). It is noted that the number of points in a particular cross-distance function depends on the vocal tract shape at a given time frame (see Story, 2007b, p. 3775). For all uses of the algorithm in this study and previously in Story (2007b), each cross-distance function was resampled with a cubic spline so that it contains 33 elements separated by equal length intervals.

This same process can be performed over consecutive XRMB time frames to generate a time-dependent cross-distance function. These will, in general, be referred to as $D_{\mathrm{VCV}}(x,t)$ where x is the distance from the lips and t is time. Shown in Fig. 3 are three-dimensional (3D) surface plots of the $D_{\mathrm{VCV}}(x,t)$'s determined for JW26's utterances [əpɑ], [ətɑ], and [əkɑ]. The three phonetic symbols on each plot indicate the approximate temporal location of the initial vowel [ə], the consonant, and the final vowel [ɑ], respectively. It is noted that, as expected, the consonant constriction is moved progressively in the posterior direction for the [p], [t], and [k] consonants. Because the cross-distance analysis only extends as far back as the posterior end of the palatal outline, the representation of the [k] is somewhat incomplete. That is, the point of maximum constriction is included, along with the vocal tract shape anterior to it, but the tract shape posterior to the constriction is not available. Although difficult to discern from the figures, it is also noted that the cross-distance does not necessarily become zero at the point of constriction. For points on the tongue this is a result of the low spatial resolution of the data. That is, if the point of contact of the tongue with an opposing surface is between two pellets, the cross-distance algorithm will not account for it. At the lips, an incomplete occlusion may appear in the cross-distance function for a bilabial closure because of potentially different degrees of compression force. Corrections for the "non-zero" occlusions based on various mathematical functions could perhaps be proposed. For the present study, however, it was decided not to alter the measured cross-distances and to assume that the spatial and temporal characteristics of the consonant constriction are accurately depicted. These three VCV surfaces served as the vocal tract shape data for which the subsequently described method was developed and tested.

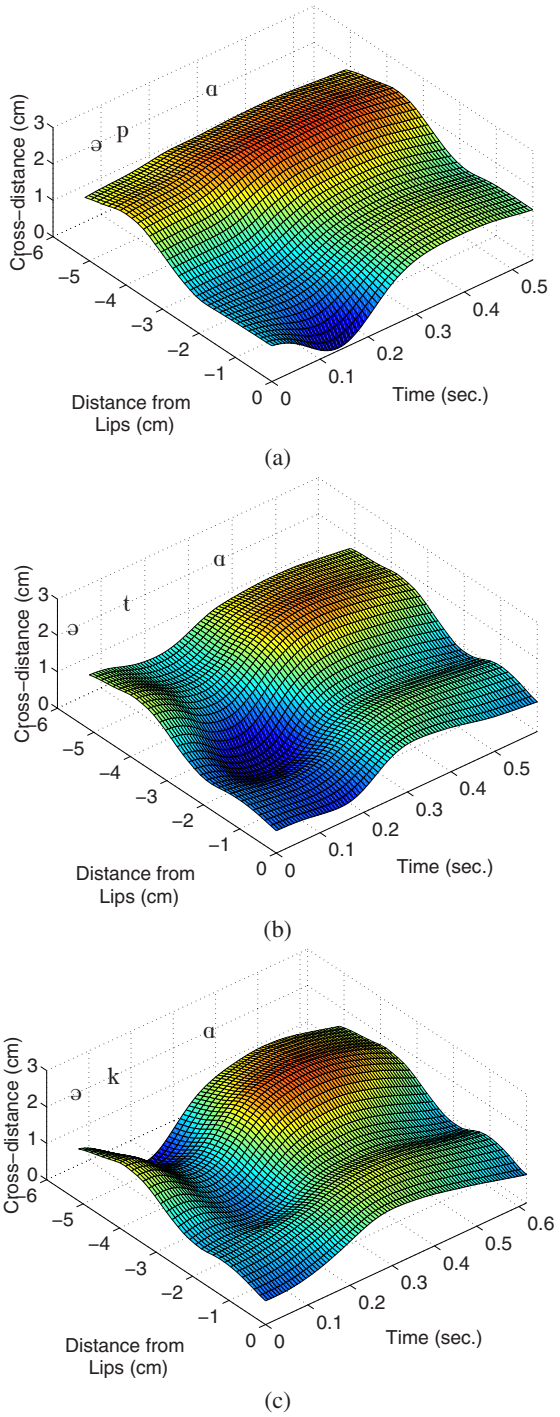


FIG. 3. (Color online) Time-varying cross-distance functions for VCVs spoken by JW26. The phonetic symbols are approximately aligned with the time frame where the respective vowels or consonants are expressed in the cross-distance function. (a) [əpɑ], (b) [ətɑ], and (c) [əkɑ].

III. RECOVERY OF THE VV SEQUENCE

The method for separating the vowel and consonant parts of a VCV cross-distance function is based on the assumption that

$$D_{\mathrm{VCV}}(x,t) = D_{\mathrm{VV}}(x,t)C(x,t) \quad \text{for } x = \Delta[1,N], \tag{1}$$

where $D_{\mathrm{VV}}(x,t)$ is the vowel-to-vowel (VV) cross-distance function that would exist in the absence of a consonant, $C(x,t)$ is a consonant superposition function that perturbs the

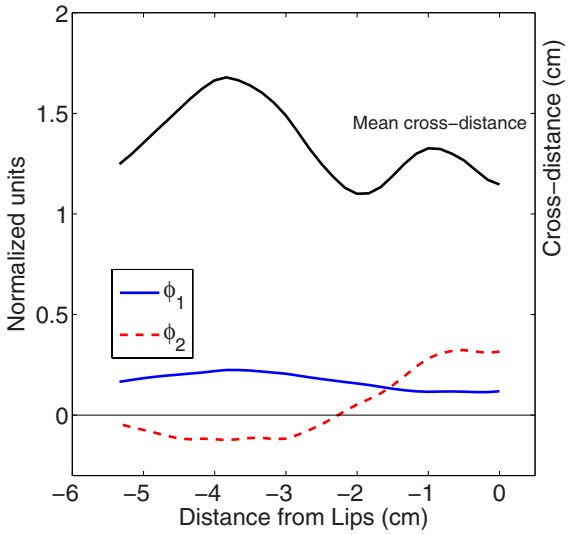


FIG. 4. (Color online) Vocal tract modes and mean tract shape for JW26 (see Story 2007b). The modes (ϕ_1 and ϕ_2) are shown in the lower part of the plot as solid and dashed lines and the mean cross-distance is shown in the upper part. Note that the axis label corresponding to the modes is on left side, whereas for the mean tract shape the label is on the right.

VV by imposing a constriction at a specific time and location, and Δ is the distance between each of the $N=33$ consecutive elements in a cross-distance function. Thus, the aim of the method is to extract a representation of both $D_{VV}(x,t)$ and $C(x,t)$ from a measured $D_{VCV}(x,t)$.

In this section, a technique is described for approximating the vowel-to-vowel sequence $D_{VV}(x,t)$ that underlies a given VCV. Using the concept of vocal tract modes, the technique effectively allows for the "recovery" of the VV sequence from the cross-distance functions generated in the previous section.

A. Vocal tract modes

It was assumed that the VV cross-distance function can be represented by time-dependent linear combinations of vocal tract modes such as those reported in Story (2007b). In that study cross-distance functions were obtained for 11 vowels from each of four speakers. Vocal tract modes were then computed, with principal component analysis, for each speaker's set of vowels. The two modes that accounted for nearly 99% of the variance in JW26's set of cross-distance functions are shown in the lower part of Fig. 4. The first mode ϕ_1 is plotted with a solid line and the second mode ϕ_2 is shown as a dashed line. The mean cross-distance function (across the eleven vowels) Ω is shown in the upper part of the plot. Although plotted in the same figure for convenience, the units for the modes and mean cross-distance are not the same, as indicated on the left and right hand sides of the plot, respectively.

Cross-distance functions can be reconstructed with high accuracy for the original 11 vowels with the relation,

$$D_V(x) = \Omega(x) + q_1\phi_1(x) + q_2\phi_2(x), \quad (2)$$

where $\phi_1(x)$ and $\phi_2(x)$ are the modes, $\Omega(x)$ is the mean cross-distance function, x is the distance from the lips, and q_1 and q_2 are weighting coefficients that generate a specific vocal tract shape.

B. Time-varying mode coefficients

The purpose of this step was to find time-dependent mode coefficients representative of the transition from initial to final vowel for each of the VCV utterances shown previously as time-varying cross-distance functions in Fig. 3. For any given VCV cross-distance function, it was hypothesized that the portions of the utterance not affected by the consonant can be fairly well approximated by a particular combination of the q_1 and q_2 coefficients of Eq. (1). It follows that those portions of the VCV that are affected by the consonant will be poorly represented by the mode coefficients. Thus, the first step in recovering the vowel and consonant components of the VCV consists of finding mode coefficients that provide a best fit to the cross-distance function at each time frame. The error between measured and reconstructed cross-distances should be low during the initial and final vowel portions and high during the consonant, and thus can provide a means of segmenting the VCV.

To determine coefficient values for the cross-distance function within each time frame of a VCV, a Nelder–Mead Simplex optimization technique (Lagarias *et al.*, 1998; the Mathworks, 2008) was used to find the $[q_1,q_2]$ coefficients that minimized the squared error between a measured cross-distance function and that constructed with Eq. (2). The minimum error within each frame can be used to generate an error function over the duration of the utterance.

Applying this process to the three $D_{VCV}(x,t)$'s in Fig. 3 produces the error functions shown in the top row of Fig. 5. Although they represent the squared error at each time frame, for purposes here, each one is normalized so that the maximum error is equal to 1.0. The points denoted with filled circles, labeled as t_o and t_f, indicate the time instants where local minima of the error functions occur, suggesting that the particular combination of coefficients determined at these points provides a good fit to the vocal tract shape. The vertical lines that pass through each of the filled circles divide the VCV into three regions. The first region ("I" in the figure), extending from the beginning of the utterance to the first local minima t_o, is the domain of the initial vowel. Between t_o and t_f is region II, a time segment where the error increases rapidly due to the influence of the consonant constriction on the underlying VV transition, and then decreases toward zero as the consonant vanishes. Region II is assumed to be a *transition* zone where the speaker simultaneously moves the vocal tract from the initial to final vowel shape and executes the onset and offset of a consonantal constriction. Region III extends from the second minimum t_f to the end of the utterance, and is assumed to be the domain of the final vowel.

Ideally, the error within regions I and III would drop to zero for any given VCV because the vowel-based modes would provide an exact match to the actual cross-distance function. Although each error function plot indicates a tendency toward zero in these regions, none of the three actually becomes zero at any point in time. For [əpɑ], there is a fairly

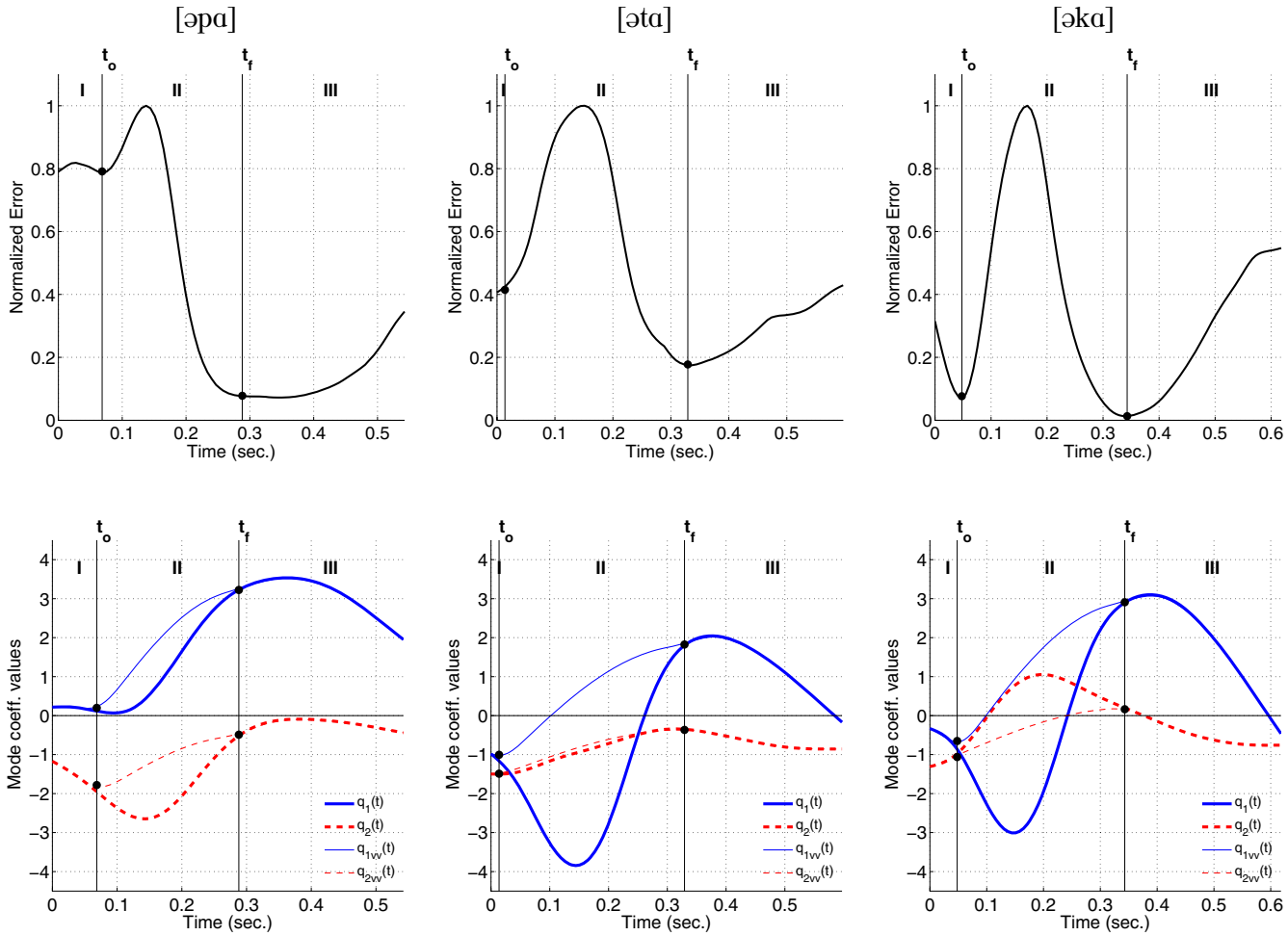


FIG. 5. (Color online) Normalized error functions (top row) and time-varying mode coefficients (bottom row) for [əpɑ] (left column), [ətɑ] (middle column), and [əkɑ] (right column). The error functions indicate the differences between the original VCV cross-distance functions and those constructed by the mode-coefficient fitting algorithm. The bottom row of plots show the time-varying mode coefficients derived from the fitting algorithm (thick lines) and the corrected versions (thin lines) based on interpolation. The regions in each plot are defined by the points, t_o and t_f, which denote the local minima on either side of the peak in the respective error function.

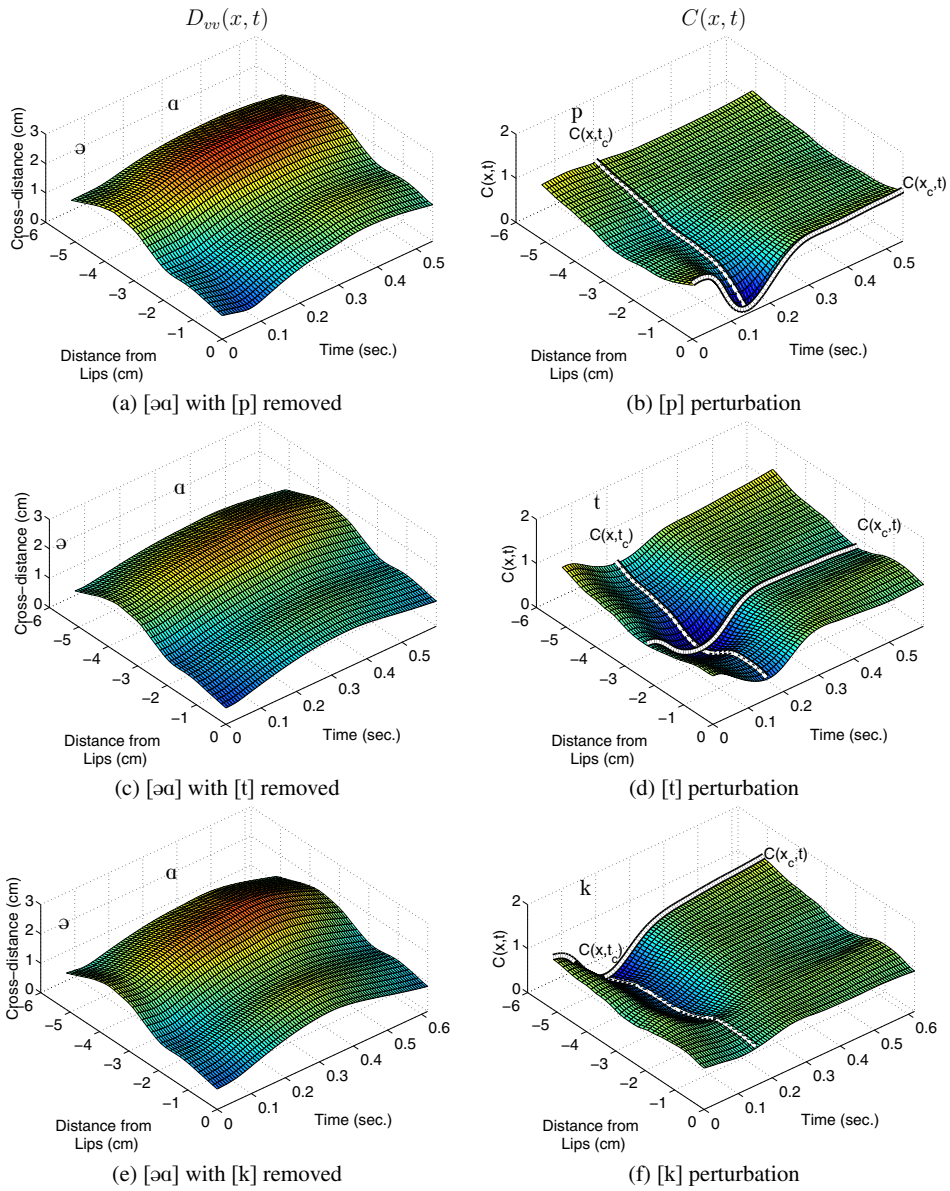
large error in region I indicating that the cross-distance functions are not well represented by the vowel-based modes, likely due to the influence of the consonant already during this period of time. For [ətɑ], a local minimum does not actually exist on the left side of the error peak, so t_o was manually chosen at a point where the slope of the error was reduced; this was done so that three regions could still be defined. The error function for [əkɑ] contains two clearly defined minima that bracket the error peak but the error does rise at the beginning of the utterance. In all three cases, the error increases toward the end of the utterance suggesting that the speaker may have been moving the vocal tract toward a less vowel-like shape that could not be well represented by the mode coefficients.

In the second row of Fig. 5 are plots of the mode coefficients as they vary over the time course of each of the three VCV utterances. The same three regions defined by the error functions are also indicated in each of these three plots. The thick lines (both solid and dashed) represent the q_1 and q_2 coefficients derived by the frame-by-frame optimization process. As would be expected because of the same target vowels, they are similar, although certainly not identical, across the three utterances in regions I and III. In region II, however, the coefficients differ across the three VCVs because the optimization process was attempting to find a vocal tract shape that would fit both the vowel shape and the given consonant constriction during this period.

To remove the effect of the consonant constriction from the temporal pattern of the mode coefficients it was assumed that, within region II, they could be replaced with an interpolation of the coefficient values at t_o (boundary of regions I and II), and their values at t_f (boundary of regions II and III). With the interpolation specified as a sinusoidal function, the $q_1(t)$ and $q_2(t)$ coefficients over the entire duration T of the utterance were replaced with

$$q_{nVV}(t) = \begin{cases} q_n(t) & \text{for } 0 \le t < t_o, \\ (q_{nf} - q_{no})\sin\left(\dfrac{2\pi(t-t_o)}{4(t_f-t_o)}\right) + q_{no} & \text{for } t_o \le t \le t_f, \\ q_n(t) & \text{for } t_f < t \le T, \end{cases} \tag{3}$$

where $n=[1,2]$, and q_{no} and q_{nf} are the coefficient values at the boundaries of region II. The $q_{nVV}(t)$ were also smoothed



(a) [əɑ] with [p] removed

(b) [p] perturbation

(c) [əɑ] with [t] removed

(d) [t] perturbation

(e) [əɑ] with [k] removed

(f) [k] perturbation

FIG. 6. (Color online) In the left column are the hypothetical VV sequences, $D_{vv}(x,t)$, recovered from the three original VCVs, and shown in the right column are the consonant perturbation functions $C(x,t)$. The product of the VV sequences with the consonant perturbations would reconstruct the original cross-distance functions. The white lines in the right column plots indicate the time frame and spatial location corresponding to the maximum constriction.

with a tenth order FIR filter that had a low-pass cutoff frequency of 10 Hz. The *thin* lines (both solid and dashed) visible in region II of the three mode coefficient plots (Fig. 5) are the result of interpolation with Eq. (3), and are representative of the underlying VV transition. Although not visible in the plots, note that $q_{1VV}(t)=q_1(t)$ and $q_{2VV}(t)=q_2(t)$ in regions I and III. The filled circle corresponding to t_o on the $q_{2VV}(t)$ trace for [əpɑ] is displaced slightly upward from the $q_2(t)$ trace. This is due to the mild filtering that is applied after the interpolation. Other interpolation schemes such as cosine, minimum jerk, or linear were also tested, but the sinusoidal function seemed to best serve the purpose of approximating the time-dependence of the coefficients during the portion of the vowel transition that is affected by the constriction. Kröger *et al.* (1995) made use of a similar sinusoid-based function for estimating movement trajectories of articulatory fleshpoints.

It is noted that division of the VCV into the three specific regions is critical for separating the vowel and consonant components. Temporal parsing by any other criterion would undermine the hypothesized representation of a VCV as a constriction superimposed on a vowel transition, and would reduce the effectiveness of the interpolation in region II to remove the influence of the consonant.

C. Reconstruction of the hypothetical VV sequence

The interpolated time-varying mode coefficients can now be used to construct a hypothetical cross-distance history over the duration of the vowel sequence with a time-dependent version of Eq. (2),

$$D_{VV}(x,t) = [\Omega(x) + q_{1VV}(t)\phi_1(x) + q_{2VV}(t)\phi_2(x)], \qquad (4)$$

in which $q_{1VV}(t)$ and $q_{2VV}(t)$ are functions of time, and $\phi_1(x)$, $\phi_2(x)$, and $\Omega(x)$ remain unchanged as functions of the distance from the lips. Time-varying vowel-to-vowel cross-distance functions were generated with the coefficients $q_{1VV}(t)$ and $q_{2VV}(t)$ from the each of the plots in the bottom row of Fig. 5, and are shown in the left column of Fig. 6. There are some subtle differences, but, as expected, each plot

shows a similar time-progression of the cross-distance from the initial [ə] to the final [ɑ] in the absence of the consonant constriction. Thus a hypothetical underlying VV transition has been recovered, in each case, from the original VCV.

IV. RECOVERY OF THE CONSONANT SUPERPOSITION FUNCTION

A. Separation of vowel and consonant portions of VCVs

With $D_{VV}(x,t)$ known for a given VCV, the time-dependence of the consonant constriction can now be recovered by rearrangement of Eq. (1) such that

$$C(x,t) = \frac{D_{VCV}(x,t)}{D_{VV}(x,t)}. \tag{5}$$

This is a ratio of the measured VCV to hypothetical VV cross-distances at every point in time and space. In regions I and III, as defined previously, this ratio should ideally be 1.0 since the cross-distances are representative of the same vowels in both the VCV and VV. The values of $C(x,t)$ in region II are expected to deviate toward 0.0 due to the influence of the consonant constriction.

With Eq. (5), $C(x,t)$ functions were calculated for the three VCVs and are shown in the right column of Fig. 6. In the initial and final vowel portions of each case, the functions are fairly flat and nearly equal to one along the posterior extent from the lips into the oral cavity. The slight deviations away from one in the vowel portions result from the inexact match of the vowel-based model to the measured cross-distance functions. Such deviations are expected based on the nonzero values present in regions I and III of the error functions (Fig. 5). In the vicinity of the consonant, each $C(x,t)$ decreases toward a minimum value near zero, revealing the onset and offset of the constriction perturbation in the absence of the vowel substrate. The temporal pattern of each constriction can be best seen along the spatial point x_c, where $C(x_c,t)$ is highlighted with a white line. The shape of $C(x,t)$ at time instant t_c represents the fully expressed consonant constriction. $C(x,t_c)$ functions are indicated on each of the plots as white lines extending from the lips back into the vocal tract.

To provide a clearer picture of the overall shape of the constrictions, all three $C(x,t_c)$ functions are replotted in Fig. 7. The constriction minima x_c, marked with filled circles, are located at the lips for [p], at −2.3 cm from the lips for [t], and at −5.1 cm for [k]. It is noted that the constriction for [k] is located at the most posterior point afforded by the XRMB data. This means that the $C(x,t_c)$ for [k] captures only the anterior portion of the constriction shape. It is also noted that even though each VCV contains a *stop* consonant, none of the three $C(x,t_c)$ functions in Fig. 7 actually becomes zero, suggesting that a complete occlusion is never attained. This is an artifact due to the low spatial resolution as discussed previously in Sec. II B.

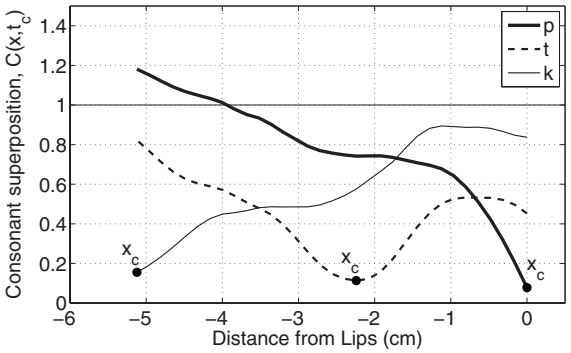


FIG. 7. Consonant perturbation (superposition) functions for [p, t, k] at the point in time representative of maximum constriction (t_c). These plots provide a clearer view of the white lines $C(x,t_c)$ shown previously in Fig. 6.

B. Time dependence of the consonant perturbation function

The operations described to this point allow for a given VCV cross-distance function to be effectively separated into the "VV" and "C" components of Eq. (1). In this section, additional operations are described for estimating the temporal activation of the consonant constriction.

The spatial configuration of the constriction is assumed to be fully expressed at time instant t_c as the function $C(x,t_c)$; these were shown previously in Fig. 7 for each of the three VCVs. This function can be represented in general as

$$C(x,t_c) = 1 - m_c(t_c)f(x) = 1 - f(x), \tag{6}$$

where $m_c(t_c)$ is a scaling function defined to be equal to one at t_c, and $f(x)$ is a spatial shaping function that, when subtracted from 1.0 at all values along the x-axis, will produce $C(x,t_c)$. At all other time instants, $C(x,t)$ is considered to be a scaled version of $C(x,t_c)$ that returns to a value of 1.0 along the entire x-axis when the consonant effect is absent.

The time course of the consonant perturbation can be determined from the variation along $C(x_c,t)$, where x_c is the constriction location. These functions were shown as white lines in the right column plots of Fig. 6. Note that each $C(x_c,t)$ tracks the temporal variation of the point of maximum constriction, and is perpendicular to the corresponding spatial variation. Each $C(x_c,t)$ can be written as

$$C(x_c,t) = 1 - m_c(t)f(x_c), \tag{7}$$

where $m_c(t)$ is the time-dependent consonant magnitude. The $m_c(t)$ function can be determined from known quantities by first substituting Eq. (7) into Eq. (1) (evaluated at x_c) to give the relation

$$D_{VCV}(x_c,t) = D_{VV}(x_c,t)[1 - m_c(t)f(x_c)]. \tag{8}$$

Rearranging Eq. (6) and evaluating at x_c yields

$$f(x_c) = 1 - C(x_c,t_c), \tag{9}$$

which can be substituted into Eq. (8). Solving for $m_c(t)$ results in

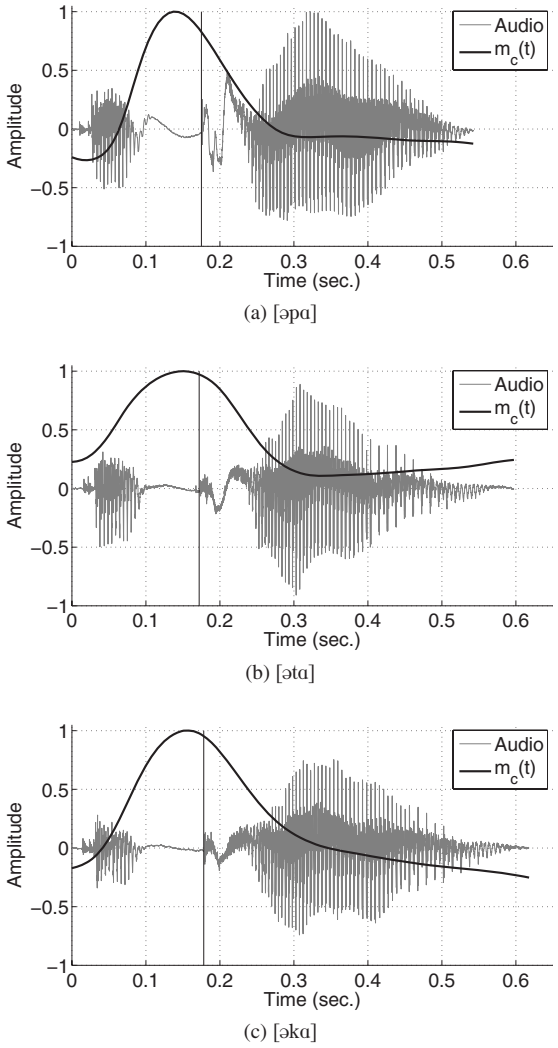


FIG. 8. Consonant magnitude functions $m_c(t)$ calculated with Eq. (10) for each of the three VCVs. Also shown in the background are the corresponding audio signals. The vertical line in each plot marks the time instant of the plosive burst.

$$m_c(t) = \frac{D_{\mathrm{VCV}}(x_c,t) - D_{\mathrm{VV}}(x_c,t)}{D_{\mathrm{VV}}(x_c,t)[C(x_c,t_c) - 1]}, \tag{10}$$

where all quantities on the right-hand side of the equation are known from previous measurements and analyses.

Plots of the magnitude $m_c(t)$ calculated with Eq. (10) for each VCV, along with the corresponding audio signal are shown in Fig. 8. In all three cases, the peak [where $m_c(t) = 1$] occurs just prior to the plosive burst as indicated by the vertical lines, and is the point at which the corresponding consonant function would be fully superimposed on the underlying VV transition. On either side of the peaks, $m_c(t)$ decreases rapidly indicating progressively less consonantal effect on the vocal tract shape. Ideally the consonant magnitude would drop to zero in these regions (i.e., no consonantal influence), but for [p] and [k], it decreases to values slightly below zero, and for [t], never reaches zero. The imperfect match of the hypothetical VV transitions to the nonconsonantal portions of the original utterances causes the numerator of Eq. (10) to be non-zero during these time periods, thus generating the positive or negative offset observed in the $m_c(t)$ functions. Nonetheless, it is apparent from these plots that the influence of the consonant begins prior to the acoustic offset of the initial vowel, and seems to subside in less than 0.1 s after the acoustic onset of the final vowel. These values are approximately representative of the three utterances spoken by the speaker JW26 and are not intended to characterize the consonants in general, but rather to show the superposition effect of the consonant on the underlying vowel substrate.

V. RECONSTRUCTION OF THE ORIGINAL VCVs

The $m_c(t)$ functions derived in the previous section can now be used to reconstruct the time-varying cross-distance functions for each of the respective VCVs. First, from Eq. (6)

$$f(x) = 1 - C(x,t_c), \tag{11}$$

which can be substituted into

$$C'(x,t) = 1 - m_c(t)f(x) = 1 - m_c(t)[1 - C(x,t_c)] \tag{12}$$

to produce a consonant superposition function. Next, Eq. (1) is rewritten with new notation as

$$D'_{\mathrm{VCV}}(x,t) = D_{\mathrm{VV}}(x,t)C'(x,t), \quad x = \Delta[1,N], \tag{13}$$

where the primes indicate an *approximation* to the previous similar quantities.

Using Eqs. (12) and (13), the $C'(x,t)$ was recovered and $D'_{\mathrm{VCV}}(x,t)$ reconstructed for [əpɑ], [ətɑ], and [əkɑ]. The resulting time-varying functions are shown as 3D plots in the left and right columns of Fig. 9. As prescribed by Eq. (13), each $D'_{\mathrm{VCV}}(x,t)$ in the right column is the product of the corresponding $C'(x,t)$ in the left column with the appropriate $D_{\mathrm{VV}}(x,t)$ shown previously in Fig. 6. Qualitatively, these plots indicate that the primary temporal and spatial features of both the original cross-distance functions (Fig. 3) and the derived consonant superposition functions (Fig. 6, right column) are maintained in the reconstructed versions, but some of the fine detail is lost.

The accuracy of each $D'_{\mathrm{VCV}}(x,t)$ relative to the original time-varying cross-distance function was assessed by calculating the rms error and correlation coefficient at each time frame. These quantities are shown in Fig. 10 as functions of time. The rms error, which provides a measure of the absolute difference in the cross-distances, is shown in the lower part of the graph and its units are denoted by the axis label on the left. With the exception of the initial 0.05 s of [əpɑ], the rms error of all three VCVs is less than 0.2 cm for the duration of the utterance. During the period of time where the consonant affected the vocal tract shape, the rms error is nearly zero. Shown in the upper part of Fig. 10, and denoted by the right axis label, is the correlation coefficient. This gives an assessment of the similarity in shape of the original and reconstructed cross-distance functions at each time frame, and is nearly 1.0 over the duration of each VCV. The exception, again, is the initial 0.05 s portion of [əpɑ] where the correlation coefficient drops to about 0.85. Taken together, these two measures suggest that the three VCVs can

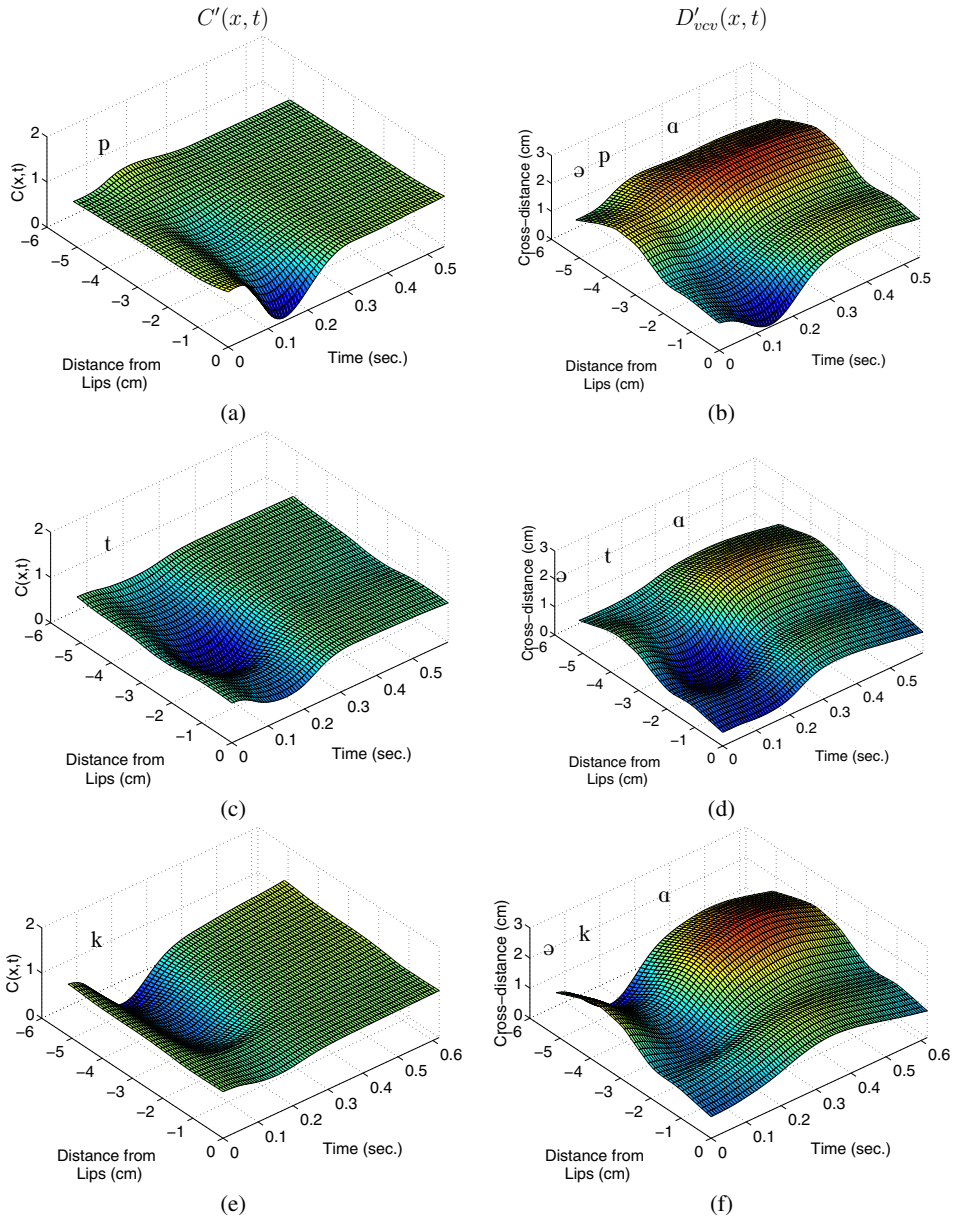


FIG. 9. (Color online) In the left column are the consonant perturbation functions $C'(x,t)$ recovered for each of the three VCVs with Eq. (12), and shown in the right column are the reconstructed VCV cross-distance functions $D'_{\mathrm{VCV}}(x,t)$.

be reasonably well represented and reconstructed by combining a separate VV sequence with a time-dependent consonant perturbation.

VI. DISCUSSION

In the discussion of his numerical model of coarticulation, Öhman (1967, p. 318) stated that "The principal value of the model…lies in its ability to summarize with a single mathematical formula the articulatory equivalent of the rather complex acoustic description of VCV coarticulation…" Perhaps the same might be said for the present model, although a few more mathematical relations than just one are required for the complete description. In particular, the ability of the VCV cross-distance model to separate the temporal and spatial contributions of vowels and consonants to the vocal tract shape may allow for insight into the planning of speech utterances. For instance, questions regarding the variability of constriction location, shape, and timing within different vowel environments may be addressed. It is also of interest to relate vocal tract shape change to acoustic output. This could be studied by comparing the structure of formant frequency contours measured for the VCV utterances with the temporal variations of the consonant magnitude and mode coefficients, similar to the idealized version of acoustic characteristics shown in Fig. 1.

A goal in developing the VCV model is to eventually use it to provide information useful for controlling an area function model of the vocal tract. Since the cross-distance model shares many features with the area function model proposed by Story (2005b), transformation of the spatial and temporal information derived for VCVs to parameters relevant for this model should be relatively straightforward. Such a transformation was demonstrated for vowel sequences (Story, 2007b) based on mode coefficients, but additional work is needed to parametrize the consonant superposition functions. Specifically, the $C(x,t_c)$ functions (Fig. 7) will need to be described in terms of constriction location (already denoted as x_c), extent of the constriction along the

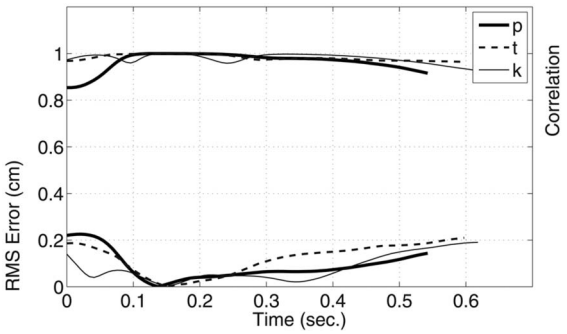


FIG. 10. Comparison of the reconstructed VCVs of Fig. 9 to the original time-varying cross-distance functions shown previously in Fig. 3. The lower part of the plot indicates the rms error between the cross-distance functions at every time frame over the course of each utterance. The upper part shows the correlation coefficients calculated at each time frame. Note the axis label for the rms error is on the left side of the plot and on the upper right side for the correlation coefficients.

vocal tract length, and degree of symmetry about the constriction location (skewness). In addition, informed techniques for correcting the $C(x,t_c)$ functions for complete occlusions need to be developed, as well as corrections for the offsets of the consonant magnitude functions $[m_c(t)]$ during the vowel-only portions of VCVs, as noted in Fig. 8.

Although speculative at this point, transformation of the VCV cross distance information to area function model parameters may allow for more realistic simulation of vocal tract shape changes along with associated acoustic output. This could aid in understanding both the acoustic and perceptual consequences of various spatial and temporal aspects of vocal tract movement. Controlling an area function with these parameters would also allow for experimenting with the temporal or spatial variability of the vowel and consonant components of an utterance. For example, the acoustic effect and perception of "sliding" the consonant magnitude function (cf. Tjaden, 1999; Löfqvist and Gracco, 1999) along the time axis relative to the vowel transition could be investigated systematically.

A key part of the VCV cross-distance model is the interpolation of the time-varying mode coefficients during the period when the consonant affects the vocal tract shape. It is this interpolation that allows for the recovery of a hypothetical VV sequence that can subsequently be used to recover the consonant superposition function. The sinusoidal function used in the model was chosen because it was found to provide a reasonable fit to the mode coefficients between the boundaries of region II. This does not mean that a sinusoidal function is necessarily representative of the temporal pattern of a complete vocal tract gesture, however. It is noted that the time points, t_o and t_f, that bound region II for each VCV are located at the onset and offset of the consonant gesture (defined by the minima in the error functions), not the beginning and end of the VV transition. Thus, the interpolation of the mode coefficients within region II is an estimate of only a portion of the entire VV transition. It is also important to note that locating t_o and t_f at the error function minima is necessary to avoid an interpolation that (1) could include remnants of the consonant constriction if Region II were reduced in duration or (2) not adequately represent the VV transition if the duration of Region II were increased.

Although the VCV cross-distance model shares many similarities with the Öhman (1967) model, a primary difference is that all of the parameters needed by the present model to decompose and reconstruct a given VCV can be obtained from analysis of that VCV alone. In Öhman's model the canonical vocal tract shape for a particular consonant and the coarticulation function could be obtained only from multiple VCVs in which that consonant was embedded within two *different* symmetric vowel contexts. Also required were canonical tract shapes for the vowels themselves. In the XRMB database, isolated VCVs were spoken only in the asymmetric [ɘɑ] context, hence, a direct comparison with the Öhman model is not currently possible unless a variety of VCVs were approximated from word and sentence level material. Additional new XRMB data collection could potentially supply a wider variety of VCV speech material for which a comparison could be made. Another difference is that the representation of velar stop consonants by the Öhman model is problematic because the location of the constriction is dependent on the vowel context, meaning that a canonical tract shape does not exist for this class of consonants. Although the VCV cross-distance approach has the problem of limited data in the velar region, there is no limitation of the model with regard to vowel-dependent constriction locations.

At this point the utility of the VCV model is admittedly limited by a number of factors. First, it has been applied to only three VCVs from one speaker. Certainly more speech material from additional speakers needs to be modeled in order to determine if the decomposition technique generalizes across a wide range of initial and final vowels, as well as across a variety of consonantal constrictions. Second, the constraint that XRMB articulatory data represent only the oral cavity does limit the degree to which the results can be extended to describe movement of the entire vocal tract. This limitation is tempered somewhat by knowing that the vocal tract modes determined from XRMB data appear to replicate those modes found for area functions based on the entire vocal tract (Story, 2007b). In addition, for most consonants other than those in the velar region, the consonant constriction is fairly well represented by the extent of the XRMB data. For the velar consonants, only a portion of the full extent of the constriction is captured by the data, and it is possible that in some cases the location of maximum constriction is not well represented. It can be noted, however, that the location and general configuration of the constrictions for each of the three consonants represented by the consonant superposition functions in Fig. 7 are similar to the vocal tract area functions reported for the same stop consonants in Story and Titze (1998). A third limitation is that the XRMB flesh points (pellets) provide a sparse spatial representation of the vocal tract in only the midsagittal plane. The low spatial resolution does not allow for a detailed representation of the air channels needed for fricative and affricate production, and midsagittal data cannot adequately depict consonants with lateral constrictions.

Despite these limitations, however, reconstruction of even a rudimentary and partial representation of the *time-varying* vocal tract shape is difficult to achieve by any means other than fleshpoint tracking [i.e., XRMB data, but also articulometer-type data (Perkell *et al.* 1992)]. New techniques in magnetic resonance imaging show promise for obtaining 3D time-dependent image sets (e.g., Takemoto *et al.* 2006), but these are currently limited to short utterances that must be repeated several hundred times. Finally, it must be pointed out that reasonable success in separating the vowel and consonant components of a VCV with the proposed model does not prove that a speaker necessarily plans an utterance by superimposing a constriction on a vowel transition. Rather, the results merely suggest it to be a possible planning paradigm that, at this stage, may provide useful information for more detailed modeling of the time-varying vocal tract shape.

VII. CONCLUSION

The aim of this study was to develop a method by which a VCV, represented as a time-dependent cross-distance function derived from XRMB articulatory data, could be separated into a vowel-to-vowel sequence and a consonant superposition function. The result is a model that represents a vowel sequence as a time-dependent perturbation of the neutral vocal tract shape governed by coefficients of the vocal tract modes. Consonants are modeled as superposition functions that can force specific portions of the tract shape to be constricted or expanded, over a specific time course. Reconstructions of three VCVs ([əpɑ], [ətɑ], and [əkɑ]) with the developed model were shown to be reasonable approximations of the original VCVs, as assessed qualitatively by visual inspection and quantitatively by calculating rms error and correlation coefficients. Although the method was developed and tested on a small data set from one female speaker, it does establish a method for future modeling of other speech material.

ACKNOWLEDGMENTS

This research was supported by NIH Grant No. R01-DC04789. The author would like to thank Kate Bunton for discussions on x-ray microbeam data and editing of early drafts of the manuscript, as well as the Associate Editor and two anonymous reviewers for their insightful comments.

[1]The model includes four hierarchical tiers where the third and fourth tiers provide time-dependent control of vocal tract length changes and nasalization, respectively. These aspects of speech production, however, are not discussed in the present study. Hence, these capabilities of the model are not reviewed.

[2]In addition to a model of the vocal tract shape, production of synthetic or simulated speech requires components for computing voice source characteristics and wave propagation. These latter two components are part of an existing speech production model (e.g., Story and Titze, 1998), but will not be discussed here.

[3]In Story (2007b) the two most anterior phantom points were determined by a correction function applied to the upper and lower lip pellet positions based on assuming contact during production of the bilabial consonant (m). In the present study, this correction function was based on lip closure during a silent non-speech segment of recorded data. This was done so that the upper and lower lip positions would indicate contact without the lip compression that is characteristic of bilabials. The vocal tract modes used in this study were also recalculated based on the new correction.

Båvegård, M. (**1995**). "Introducing a parametric consonantal model to the articulatory speech synthesizer," in Proceedings Eurospeech 95, Madrid, Spain, pp. 1857–1860.
Browman, C., and Goldstein, L. (**1990**). "Gestural specification using dynamically-defined articulatory structures," J. Phonetics **18**, 299–320.
Byrd, D. (**1996**). "Influences on articulatory timing in consonant sequences," J. Phonetics **24**, 209–244.
Carré, R., and Chennoukh, S. (**1995**). "Vowel-consonant-vowel modeling by superposition of consonant closure on vowel-to-vowel gestures," J. Phonetics **23**, 231–241.
Fowler, C. A. (**1980**). "Coarticulation and theories of extrinsic timing," J. Phonetics **8**, 113–133.
Fowler, C. A., and Saltzman, E. (**1993**). "Coordination and coarticulation in speech production," Lang Speech **36**, 171–195.
Gracco, V. L. (**1992**). "Characteristics of speech as a motor control system," Haskins Labs. Stat. Rep. on speech Res., SR-109/110, 13-26.
Ichikawa, A., and Nakata, K. (**1968**). "Speech synthesis by rule," Reports of the Sixth International Congress on Acoustics, edited by Y. Kohasi (International Council of Scientific Unions, Tokyo), pp. 171–1744.
Kent, R. D., and Minifie, F. D. (**1977**). "Coarticulation in recent speech production models," J. Phonetics **5**, 115–133.
Kozhevnikov, V. A., and Chistovich, L. A. (**1965**). "Speech: Articulation and perception (trans. US Dept. of Commerce, Clearing House for Federal Scientific and Technical Information)," Joint Publications Research Service, Washington, DC, No. 30, p. 543.
Kröger, B. J., Schröder, G., and Opgen-Rhein, C. (**1995**). "A gesture-based dynamic model describing articulatory movement data," J. Acoust. Soc. Am. **98**, 1878–1889.
Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E. (**1998**). "Convergence properties of the Nelder–Mead Simplex method in low dimensions," SIAM J. Optim. **9**, 112–147.
Löfqvist, A., and Gracco, V. L. (**1999**). "Interarticulator programming in VCV sequences: Lip and tongue movements," J. Acoust. Soc. Am. **105**, 1864–1876.
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.
Nakata, K., and Mitsuoka, T. (**1965**). "Phonemic transformation and control aspects of synthesis of connected speech," J. Radio Res. Labs. **12**, 171–186.
Öhman, S. E. G. (**1966**). "Coarticulation in VCV utterances: Spectrographic measurements," J. Acoust. Soc. Am. **39**, 151–168.
Öhman, S. E. G. (**1967**). "Numerical model of coarticulation," J. Acoust. Soc. Am. **41**, 310–320.
Perkell, J. (**1969**). *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study* (MIT, Cambridge, MA).
Perkell, J., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabicta, I., and Jackson, M. T. T. (**1992**). "Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements," J. Acoust. Soc. Am. **92**, 3078–3096.
Story, B. H. (**2005a**). "A parametric model of the vocal tract area function for vowel and consonant simulation," J. Acoust. Soc. Am. **117**, 3231–3254.
Story, B. H. (**2005b**). "Synergistic modes of vocal tract articulation for American English vowels," J. Acoust. Soc. Am. **118**, 3834–3859.
Story, B. H. (**2007a**). "A comparison of vocal tract perturbation patterns based on statistical and acoustic considerations," J. Acoust. Soc. Am. **122**, EL107–EL114.
Story, B. H. (**2007b**). "Time-dependence of vocal tract modes during production of vowels and vowel sequences," J. Acoust. Soc. Am. **121**, 3770–3789.
Story, B. H., and Titze, I. R. (**1998**). "Parameterization of vocal tract area functions by empirical orthogonal modes," J. Phonetics **26**, 223–260.
Takemoto, H., Honda, K., Masaki, S., Shimada, Y., and Fujimoto, I. (**2006**). "Measurement of temporal changes in vocal tract area function from 3D cine-MRI data," J. Acoust. Soc. Am. **119**, 1037–1049.
Tjaden, K. (**1999**). "Can a model of overlapping gestures account for scanning speech patterns?," J. Speech Lang. Hear. Res. **42**, 604–617.
The Mathworks, MATLAB, Version 7.6.0.324 (R2008a).
Westbury, J. R. (**1994**). X-Ray Microbeam Speech Production Database User's Handbook, (version 1.0)(UW-Madison).

The sensitivity of auditory-motor representations to subtle changes in auditory feedback while singing

Dwayne Keough
Department of Psychology, Wilfrid Laurier University, Waterloo, Ontario N2L 3C5, Canada

Jeffery A. Jones[a)]
Department of Psychology and Centre for Cognitive Neuroscience, Wilfrid Laurier University, Waterloo, Ontario N2L 3C5, Canada

(Received 19 August 2008; revised 13 May 2009; accepted 13 May 2009)

Singing requires accurate control of the fundamental frequency (F0) of the voice. This study examined trained singers' and untrained singers' (nonsingers') sensitivity to subtle manipulations in auditory feedback and the subsequent effect on the mapping between F0 feedback and vocal control. Participants produced the consonant-vowel /ta/ while receiving auditory feedback that was shifted up and down in frequency. Results showed that singers and nonsingers compensated to a similar degree when presented with frequency-altered feedback (FAF); however, singers' F0 values were consistently closer to the intended pitch target. Moreover, singers initiated their compensatory responses when auditory feedback was shifted up or down 6 cents or more, compared to nonsingers who began compensating when feedback was shifted up 26 cents and down 22 cents. Additionally, examination of the first 50 ms of vocalization indicated that participants commenced subsequent vocal utterances, during FAF, near the F0 value on previous shift trials. Interestingly, nonsingers commenced F0 productions below the pitch target and increased their F0 until they matched the note. Thus, singers and nonsingers rely on an internal model to regulate voice F0, but singers' models appear to be more sensitive in response to subtle discrepancies in auditory feedback.
 [DOI: 10.1121/1.3158600]

PACS number(s): 43.70.Mn, 43.70.Bk, 43.70.Gr [AL] Pages: 837–846

I. INTRODUCTION

The role of auditory feedback during vocalization has been a topic of much scientific inquiry. Both speech and singing are debatably the most complex motor actions humans are capable of producing. In order to produce a word, or to sing a musical note, one must possess strict control over respiratory muscles in addition to control over intrinsic and extrinsic laryngeal muscles. This control is achieved by an intricate network of cortical and brainstem areas dependent upon auditory (Sapir *et al.*, 1983; Larson *et al.*, 2008) and proprioceptive (Kirchner and Wyke, 1965; Wyke, 1974; Yoshida *et al.*, 1989) reflex mechanisms. However, nonreflexive systems that utilize auditory feedback also contribute greatly to the development of speech in children (e.g., Oller and Eilers, 1988).

Postlingually, auditory feedback also appears to be monitored and used during ongoing speech. For instance, delaying auditory feedback results in disruptions in the quality of vocal productions (Yates, 1963). Moreover, altering auditory feedback typically elicits compensatory responses in speakers' ongoing vocal productions. Increases in masking noise and decreases in side-tone amplitude (Bauer *et al.*, 2006; Lane and Tranel, 1971), shifts in formant frequencies (Houde and Jordan, 1998; Purcell and Munhall, 2006), and the fundamental frequency (F0) (Burnett *et al.*, 1997; Elman, 1981; Jones and Munhall, 2000; Kawahara, 1998) all cause vocal responses that oppose the respective manipulations. The purpose of the present study is to further our understanding of the role of auditory feedback for the control of F0.

Frequency-altered feedback (FAF) is one paradigm that has been used to examine the importance of auditory feedback for the regulation of voice F0. In a typical FAF study, participants receive auditory feedback regarding their pitch that is higher or lower than their actual vocal productions. Responses to upward or downward perturbations generally result in decreases or increases in F0, respectively. Interestingly, the majority of FAF studies (e.g., Burnett *et al.*, 1997; Burnett *et al.*, 1998; Burnett and Larson, 2002; Elman, 1981; Jones and Munhall, 2000, 2002, 2005; Natke *et al.*, 2003; Toyomura *et al.*, 2007) had participants produce a relative target, but not match a specific frequency, as one must do while singing.

In those studies, most participants compensated for the discrepancy detected between perception and production, although not perfectly (e.g., Burnett *et al.*, 1997; Donath *et al.*, 2002; Larson, 1998; Larson *et al.*, 2000; Natke *et al.*, 2003). On average, manipulations varying from ±100 to 600 cents had a response magnitude of approximately 50 cents (regardless of the direction of manipulation) but ranged from 15–65 cents (cent is a logarithmic unit used to measure small intervals between different frequencies, where 100 cents is equal to 1 semitone). Liu and Larson (2007) observed complete compensation for small perturbations (10 cents), suggesting that the role of the auditory-vocal system is to correct for

a)Author to whom correspondence should be addressed. Electronic mail: jjones@wlu.ca

small errors rather than larger F0 deviations. However, it is worth noting that the response magnitudes obtained from previous studies (e.g., Burnett *et al.*, 1997; Donath *et al.*, 2002; Larson, 1998; Larson *et al.*, 2000; Liu and Larson, 2007) were measured by the random presentation of auditory perturbations, whereas the current study assessed sensorimotor adaptation through the presentation of predictable changes in voice F0.

Singing offers researchers a unique window into the study of F0 control. Singing involves producing a succession of musical sounds at a particular (absolute) frequency (e.g., 440 Hz, A4). Thus, singers must maintain accurate vocal control over their F0 to hit the desired notes. Sundberg (1987) (p. 177) reported that if the actual vocal production deviates from the intended target, trained singers are able to compensate and match the tone (A4, 440 Hz) with an accuracy of less than 1 Hz. However, few studies have directly examined the role of auditory feedback while singing (Burnett *et al.*, 1997; Jones and Keough, 2008; Natke *et al.*, 2003; Zarate and Zatorre, 2005, 2008). The results obtained during singing studies that have used FAF (e.g., Burnett *et al.*, 1997; Natke *et al.*, 2003) indicate that the compensation is comparable to those achieved during speech production. The compensatory responses observed in trials of altered feedback indicate that voice production is regulated in a closed-loop fashion (Fairbanks, 1954; Hain *et al.*, 2000; Larson *et al.*, 2008; Lee, 1950). However, laryngeal structures and vocal fold stiffness are set in place prior to vocal onset (e.g., Watts *et al.*, 2003), suggesting a role for open-loop motor planning.

In the case of singing, when learning to produce a target note, stronger reliance on auditory feedback during the initial acquisition stages may be required in order to establish sensorimotor representations that will guide future vocal productions. This idea is based on the premise that the feedback one receives while learning to perform a specific task creates an integrated sensorimotor representation that is directly related to the task (Proteau *et al.*, 1987; Proteau *et al.*, 1992). For instance, Finney and Palmer (2003) demonstrated that the availability of auditory feedback while pianists learned a piece of music significantly improved their ability to play the piece from memory during later recall. However, when pianists were asked to perform well-rehearsed musical sequences from memory, the removal of auditory feedback did not affect musical performance (Finney and Palmer, 2003).

Thus, over time, it is hypothesized that an internal representation is formed or "internal model" that stores the relationships between the motor commands, environment, and sensory feedback for their production (Proteau *et al.*, 1987; Proteau *et al.*, 1992). As a result, pianists may be capable of performing a well-rehearsed piece from memory without auditory feedback because they are relying on internal motor commands that correspond to the musical composition. The same reliance on an internal motor plan may also exist for singing a well-rehearsed piece. For instance, it is possible that vocal productions are initiated based on an increased reliance on the motor plan established for vocalization.

Jones and Keough (2008) investigated whether trained singers, given their extensive training, rely more on a well-established internal representation than nonsingers while singing. Indeed, identifying differences in F0 control while participants receive subtle manipulations in auditory feedback should elucidate the sensitivity of the underlying mechanisms regulating F0 control. Jones and Keough (2008) found that nonsingers initially compensated to a greater degree than singers when exposed to FAF (−100 cents). However, after this brief exposure to FAF, singers' F0 values were higher than their base line F0 values when they heard their feedback returned to normal. In contrast, no differences were found between the F0 values during the base line and test trials for nonsingers. These effects persisted when singers were required to produce a note other than the one they sang during the altered feedback trials. The results imply that singers rely more on an internal model to regulate their F0 productions during singing than nonsingers.

The present study was designed to investigate the sensitivity of singers' and nonsingers' internal representations to subthreshold (2 cent increments) manipulations in auditory feedback. The data obtained by Jones and Keough (2008) suggested that trained singers compensated less for larger discrepancies between perception and production than nonsingers and instead relied more on their internal models. Nonsingers compensated immediately after exposure to altered feedback of 100 cents, whereas singers required several trials to modify their F0 (Jones and Keough, 2008). Moreover, Zarate and Zatorre (2005, 2008) found that singers were able to successfully ignore the altered auditory feedback (by presumably relying on an internal model) they received and to continue to produce the targets at the desired frequency when instructed to do so. Conversely, nonsingers' vocal productions suggested that they were unable to ignore the altered feedback and as a result adjusted their F0 to compensate for the manipulation (Zarate and Zatorre, 2005, 2008).

Jones and Keough (2008) found that singers initially responded less to large (100 cent) shifts in auditory feedback than nonsingers, who exhibited near perfect levels of compensation almost immediately. However, it remains unknown whether similar F0 values for singers and nonsingers would be obtained using small manipulations in auditory feedback (i.e., 2 cent increments up to 1 semitone, 100 cents). Given the paucity of data, the authors hypothesized that singers and nonsingers would compensate to a similar degree to the altered feedback. However, they expected that singers, due to their extensive training and experience, would compensate for the discrepancy in perception and production more efficiently than nonsingers. That is, the mechanisms that regulate F0 control in singers would be more sensitive to subtle changes in FAF. As a result, singers will reproduce the target notes more accurately than nonsingers while receiving FAF.

Of particular interest was determining when singers' and nonsingers' compensatory responses to small manipulations in auditory feedback would occur. Determining when both groups initiate compensatory responses may also provide an index of how sensitive the underlying mechanisms of the internal representations are to perturbations. Recently, the just-noticeable difference of a single fundamental frequency (120 Hz) to natural sound stimuli (e.g., lowest tone on a violin pitch shifted to 120 Hz; German vowels /i:/ and /a:/

and Italian vowels /i/ and /a/) was shown to be between 4 and 98 cents (Pape and Mooshammer, 2006). Furthermore, Loui *et al.* (2008) found that participants without speech or hearing disorders, and with no formal music training, had a perceptual and production threshold (approximately 2.0 and 2.5 Hz, respectively) that was significantly better than the threshold for amusical ("tone-deaf") participants (approximately 36.2 and 12.3 Hz, respectively). The authors hypothesized that singers would adapt to altered feedback earlier than nonsingers due to their extensive vocal experience achieving pitch targets.

Previous studies using the FAF paradigm have found evidence for sensorimotor adaptation by measuring the mean F0 after feedback was returned to normal (Jones and Munhall, 2000, 2002, 2005; Jones and Keough, 2008). These aftereffects can be observed following exposure to a single trial of FAF (Natke *et al.*, 2003) in nonsingers. To track sensitivity to FAF in this experiment, the authors assessed sensorimotor adaptation by measuring F0 at vocal onset while exposing participants to predictable changes in auditory feedback (increases or decreases in frequency by 2 cents on each successive trial). Using this measure, aftereffects will be evident when voice F0 values start at or near F0 values observed at the end of the previous utterance.

If internal models are continuously updated, then vocalization onset should progressively become lower and higher when participants receive FAF that is either increasing or decreasing, respectively. Thus, when participants receive unaltered auditory feedback at the end of the experiment, adaptation will be demonstrated if their F0 values begin near their F0 values from the final FAF trial. As a result, if singers rely more on an internal representation to control voice F0, then aftereffects should be more pronounced in singers than in nonsingers. Moreover, adaptation effects should generalize to a greater degree in singers than in nonsingers when asked to produce a different note (A4 and F4) with unaltered feedback following FAF trials.

II. METHODS

A. Participants

Twenty Wilfrid Laurier University students (all women) whose native tongue was North American English participated. Although there is no evidence to suggest that a gender difference exists in response to FAF, men were excluded so that all participants could adequately sing the same target notes. Of the 20 participants, 10 were trained singers (mean musical training was approximately 12 years) recruited from the faculty of music (vocal majors) at Wilfrid Laurier University. None of the trained singers reported having "perfect" pitch. The remaining ten participants were considered to be nonsingers, as none possessed any form of vocal training or ongoing participation in formal singing. All participants passed a bilateral hearing test at 20 dB for frequencies at 250, 500, 1000, 2000, 4000, and 8000 Hz and received financial compensation for their time and gave informed consent. The Wilfrid Laurier University Research Ethics Committee approved the procedures.

B. Apparatus

1. Participant recording sessions

Participants were situated in a double-walled sound attenuated booth (Industrial Acoustic Co., model 1601-01) and were fitted with headphones (Sennheiser HD 280 Pro) and a condenser microphone (Countryman Isomax E6 Omnidirectional Microphone), which was approximately 3 cm from their mouth. Multitalker babble noise (20 speakers simultaneously reading different passages; Auditec, St. Louis, MO) was presented at 80 dB SPL (sound pressure level) to limit natural acoustic feedback. Multitalker babble is the sound of 20 young adults simultaneously reading different passages, which is unintelligible to the listener. The target notes consisted of a female voice singing the consonant-vowel /ta/ that was presented at either 349, 392, or 440 Hz, respectively. Microphone signals were sent to a signal processor (VoiceOne 2.0, TC Helicon) that manipulated auditory feedback. The altered feedback was then mixed (Mackie ONYX 1640) with the multitalker babble and subsequently sent to the participant. Vocal productions were digitized at 44.1 kHz for future analysis.

2. Target stimuli recording

The stimuli used were the same as in Jones and Keough (2008). A trained singer produced the respective targets, F4, G4, and A4, which were processed using the speech modification algorithm STRAIGHT (speech transformation and representation using the adaptive interpolation of weighted spectrum; Kawahara *et al.*, 1999) to ensure that each target was exactly 349, 392 or 440 Hz.

C. Procedure

Participants matched a musical target over 210 trials, which were divided into three blocks of 70 trials. Each block consisted of 10 base line (only the last five trials were subjected to statistical analysis), 50 shift, and 10 test trials. During one block, participants reproduced the musical target G4 on all trials. Thus, participants received unaltered auditory feedback during the ten base line trials, followed by 50 FAF trials (2 cent increments to 100 cents, 1 semitone), and then ten more unaltered feedback trials. In two other blocks, participants reproduced either F4 or A4 for the ten base line and ten test trials while singing G4 during the FAF trials. Thus, participants only sang G4 during FAF trials. The other target notes, A4 and F4, were used to test whether participants exhibited sensorimotor adaptation when emulating notes that were never manipulated. Thus, the three block participants experienced were AGA, FGF, and GGG, where the first letter denotes the base line target (trials 1–10), the middle letter denotes the shifted target (trials 11–60), and the final letter denotes the test value (trials 61–70). The three blocks of trials were counterbalanced across participants.

On the first day of testing, participants produced the target on the three blocks (AGA, FGF, and GGG) of 70 trials and received auditory feedback that was shifted either upward or downward in 2 cent increments up to 100 or −100 cents, respectively, during the shift trials. Note that auditory feedback was shifted from the beginning of each utterance

until the end of their vocal productions. On a subsequent day, participants produced the remaining three blocks of 70 trials while receiving altered feedback that was shifted in the opposite direction. Thus, participants were required to produce the target for a total of six blocks of 70 trials that were counterbalanced over two days. Although aftereffects persisted for the duration of test trials in a previous study (Jones and Keough, 2008), there is no evidence in the literature to suggest that aftereffects continue to occur when tested on subsequent days. Individual trials commenced with the presentation of multitalker babble for 1000 ms followed by the target stimulus, which lasted 2000 ms in duration. Lastly, the multitalker babble was presented again for 4000 ms.

The babble was presented prior to the target to inform participants that the target note would be forthcoming. During the second presentation of the babble, participants were instructed to sing the target note as accurately as possible in pitch and duration (hold their pitch constant for approximately 2000 ms). Trials were initiated and controlled by a computer, and participants' vocal productions were recorded on a Macintosh G5 computer (TRACKTION v1, Woodinville, WA). F0 values for each vocal production were calculated, during offline analyses, using an autocorrelation algorithm included in the PRAAT program (Boersma, 2001). F0 values were normalized to each target note (F4, G4, or A4) by calculating the appropriate cent values using the following formula:

$$\text{Cents} = 100(12 \log_2 F/B)$$

where F is the F0 value in hertz and B is frequency of the target pitch participants were instructed to sing (349, 392, or 440 Hz).

The mean F0 values for singers and nonsingers during FAF trials (11–60) were calculated for each condition (AGA, GGG, and FGF) and are displayed in Figs. 1(a) and 1(b) (gray lines), respectively. Data for one nonsinger were removed from the statistical analyses as she exhibited poor F0 control during testing (mean F0 was 346 and 244 cents for the shifted up and down conditions, respectively). Only the first 1500 ms for each vocal production was analyzed because previous research has found that compensatory responses to FAF occur between 130 and 500 ms after perturbation onset (Burnett *et al.*, 1997; Burnett *et al.*, 1998; Jones and Munhall, 2002). The F0 values from the pitch shift up and down AGA, FGF, and GGG blocks were analyzed together and were broken into five different blocks of trials within each condition: shift trials (11–20, 21–30, 31–40, 41–50, and 51–60). Furthermore, the first five trials and the last five trials of each block were averaged and divided into early and late phases, respectively. Thus, a multivariate analysis of variance (MANOVA) was carried out on the mean F0 values with 2(experience:singer and nonsinger)×2(pitch shift:up and down)×5(block)×2(phase: early and late) as factors. Newman–Keuls' test was used for *post hoc* tests with an alpha level of 0.05 used for all statistical tests.

Also, the authors wanted to determine when singers' and nonsingers' mean F0 values during altered feedback trials were significantly different from base line F0 values. This test would indicate when compensation occurred in the subtle manipulations in feedback. Furthermore, the authors examined the first test trial immediately after participants received altered feedback to determine if sensorimotor adaptation, in the form of aftereffects, occurred. In order to assess compensatory responses and aftereffects, multiple *t*-tests were performed on the mean F0 value for the last five base line trials compared with the mean F0 values on altered feedback trials and the first test trial during shifted up and down conditions.

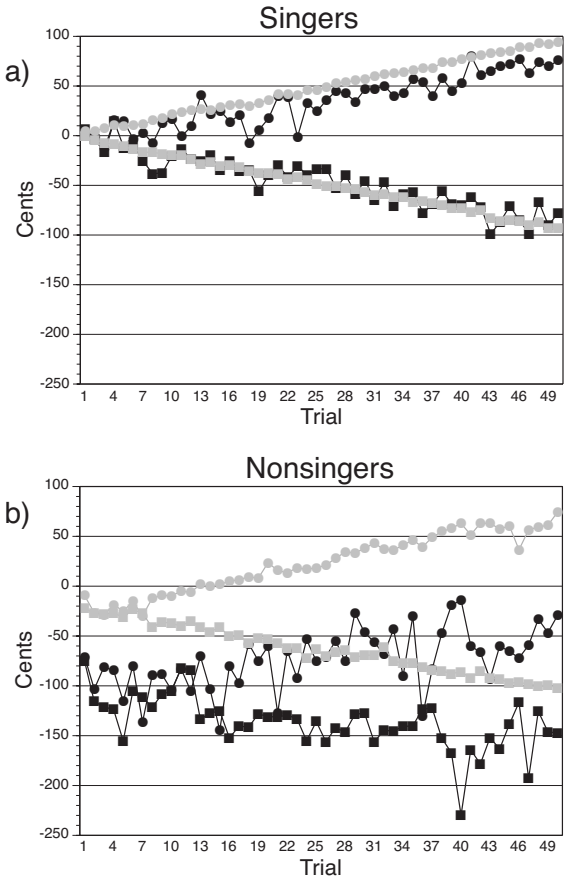


FIG. 1. (a) Singers' and (b) nonsingers' mean F0 (gray symbols) over a whole utterance and median F0 (black symbols) for the first 50 ms after vocal onset during FAF trials (11–60) for each condition (AGA, GGG, and FGF). Each data point represents the mean compensatory response to each incremental manipulation (three trials per shift value) averaged in blocks of 10 cent shifts during both upward and downward manipulations. Participants were required to match the musical target note G4 (392 Hz). The circles depict utterances that participants heard shifted downward; the squares depict utterances that participants heard shifted upward.

Additionally, the median F0 value for the first 50 ms [see Figs. 1(a) and 1(b), black lines] of each utterance during the AGA, FGF, and GGG conditions was calculated for both singers and nonsingers across shifted trials. Determining whether differences occur between the initial 50 ms of initial shift trials and the later shifted trials indicated whether sensorimotor adaptation occurred during successive FAF trials. Furthermore, these median values were subtracted from the mean F0 over the entire utterance to determine the difference between where participants initiated vocal pitch and where they maintained their F0 values while singing. This test identified whether singers' and nonsingers' F0 productions were initiated at the desired target frequency or whether they per-

formed a searching strategy, increasing or decreasing their F0 to match the target. The median values for the first 50 ms and the median less the mean values were then categorized into five blocks of trials and further divided into two phases, in the same fashion as the mean F0 values. Thus, two separate MANOVAs were performed on each of the aforementioned values (median and median minus mean) during the shifted up and down AGA, FGF, and GGG conditions with 2(experience:singer and nonsinger)×2(pitch shift:up and down)×5(block)×2(phase:early and late) as factors.

Finally, in order to elucidate when or if sensorimotor adaptation commences during FAF trials and whether aftereffects persist during test trials, multiple *t*-tests (Bonferroni corrected) were conducted on the median F0 values for the first 50 ms during both the shifted up and down conditions. The average of the median F0 values for the last five base line trials was compared with the median F0 values of all shift trials and the first test trial following exposure to FAF.

III. RESULTS

The MANOVA carried out on the mean F0 values with 2(experience:singer and nonsinger)×2(pitch shift:up and down)×5(block)×2(phase:early and late) as factors revealed a main effect of experience and pitch shift, $F(1,55)=12.59, p<0.05$ and $F(1,55)=634.73, p<0.05$, respectively. Nonsingers' mean F0 values during FAF trials were found to be significantly lower than the pitch target (392 Hz, G4) compared to the mean F0 values of singers [see Figs. 1(a) and 1(b)]. Also, the mean F0 values during the pitch shift up condition were found to be lower than the F0 values during the pitch shift down condition. A significant two-way interaction was found between pitch shift and block, $F(4,220)=423.40, p<0.05$. *Post hoc* analysis revealed that the first block (trials 11–20) of mean F0 values during the pitch shift up condition and the pitch shift down condition was significantly different (progressively lower and higher during pitch shift up and down conditions, respectively) from all other blocks of pitch shift F0 values obtained during FAF trials (21–60) ($p<0.05$) [see Figs. 1(a) and 1(b)]. Note that each data point in Fig. 1 represents the mean compensatory response to each incremental manipulation (three trials per shift value; one from each of AGA, GGG, and FGF conditions) averaged in blocks of 10 cent shifts during both upward and downward manipulations. This pattern demonstrates that participants were compensating more during later trials to increasing and decreasing shifts in auditory feedback. Interestingly, this suggests that both singers and nonsingers initiate compensatory responses quite early to subtle changes in auditory feedback. Furthermore, the mean F0 values obtained for each pitch shift block during the shift up condition were significantly lower than the mean F0 values for each block in the shift down condition ($p<0.05$).

A two-way interaction between pitch shift and phase was also found to be significant, $F(1,55)=121.05, p<0.05$. *Post hoc* analysis indicated that the mean F0 values during the first phase of the shift up condition and the shift down condition were significantly different from the mean F0 values during the second phase of the shift up and down conditions, respectively ($p<0.05$). This implies that as pitch shifts progressively increased or decreased on FAF trials participants correspondingly adjusted their F0 to continue to produce the target notes accurately, and given that the late phase shift trials were larger than early phase shift trials it is not surprising that the mean F0 values were found to differ, on average, between the two phases. Also, the F0 values during the first and second phases of the shift up condition differed significantly from the F0 values of the first and second phases of the shift down condition ($p<0.05$). No other significant main effects or interactions were observed.

Multiple *t*-tests (uncorrected) were carried out on the mean F0 during the shifted up and down conditions for singers and nonsingers. When three consecutive significant differences were found during the shifted trials, the first significant response from base line was used to indicate the initial compensatory response. During the shift up trials, when the mean of the last five base line trials was compared with the mean of the shift trials it was found that singers initiated compensatory responses on the third (6 cents) shift trial, $t(60)=-2.89, p<0.05$. The same initial response was also observed during the shift down condition, $t(60)=2.67, p<0.05$. Moreover, singers' F0 values on shift up and down trials remained significantly different from their base line F0 values. Thus, singers' F0 values on shifted trials were significantly different from the average of the last five base line trials on 48 of 50 trials during both shift up and down conditions.

On the other hand, during the shift up condition, nonsingers were found to initiate compensatory responses on shift trial 13 (26 cents), $t(60)=-3.68, p<0.05$, with the remaining F0 values remaining significantly different from the base line F0 values. The only difference between the shift up and down conditions for nonsingers was that they initiated compensatory responses two trials earlier during the down condition, at shift trial 11 (22 cents), $t(60)=3.12, p<0.05$. Thus, nonsingers' F0 values were different from base line on 38/50 and 40/50 during the shift up and down conditions, respectively.

The authors were also interested in determining whether compensating for FAF would result in aftereffects (for the GGG condition) when auditory feedback was returned to normal and whether these aftereffects would generalize (for the AGA and FGF conditions) to a note other than the one participants received during testing. Thus, *t*-tests were carried out on the average of the mean F0 values of the last five base line trials for the AGA, FGF, and GGG conditions and the first test trial (when auditory feedback was returned to normal following FAF trials) of each respective condition [see Figs. 2(a) and 2(b)]. Nonsingers' average base line F0 values for the GGG trials were significantly different from the F0 values for the initial test trial for both shift down and up conditions, $t(8)=-2.90, p<0.05$ and $t(8)=2.30, p<0.05$, respectively. Moreover, singers' base line mean F0 values for both shift down and up conditions were also significantly different from their initial mean F0 values during testing, $t(9)=-5.39, p<0.05$ and $t(9)=4.49, p<0.05$, respectively. Thus, both singers and nonsingers exhibited aftereffects following FAF trials. That is, their mean F0 values were above

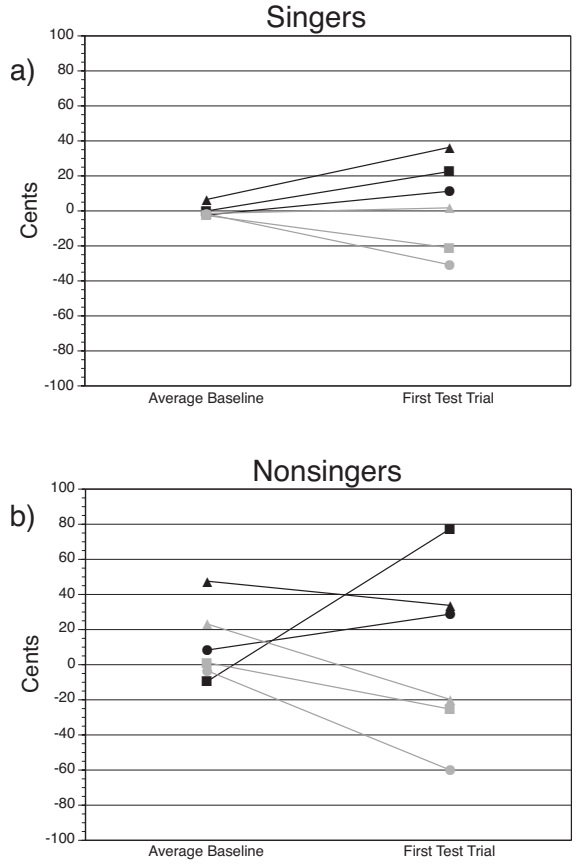


FIG. 2. The mean F0 of utterances produced by (a) singers and (b) nonsingers, averaged over the last five base line trials (prior to receiving FAF), and the mean F0 for the first test trial utterance following FAF. The triangles represent F0 values for the target note F4 (349 Hz), the squares represent F0 values for the target note G4 (392 Hz), and the circles represent F0 values for the target note A4 (440 Hz). Gray symbols represent F0 values obtained during the upward shift condition, and the black symbols represent F0 values obtained during the downward shift condition.

and below the target F0 when auditory feedback was returned to normal following exposure to upward and downward FAFs.

When participants sang a different note following FAF trials (either F4 or A4), nonsingers' mean F0 values were significantly different from their average base line F0 values for the shifted up AGA condition, $t(8)=2.58, p<0.05$. On the other hand, singers' mean base line F0 values were significantly different from their mean test F0 values for both the AGA and FGF shifted down conditions, $t(9)=-2.35, p<0.05$ and $-3.72, p<0.05$, respectively, and the AGA shifted up condition, $t(9)=5.51, p<0.05$. Overall, both groups exhibited aftereffects that generalized to at least one note other than the one produced during FAF trials. However, when the alpha level was corrected (Bonferroni) for multiple *t*-tests, nonsingers' mean F0 base line values failed to remain statistically significant when compared to the initial test trial for all conditions. Singers' aftereffects during GGG trials remained significantly different during both shifted down and up conditions, but the only generalization effect to remain significant were the differences in F0 values obtained during the shifted up AGA base line and test conditions.

A MANOVA was performed on the median 50 ms F0 values (the median F0 value within the first 50 ms after vocal onset) during the shifted up and down AGA, FGF, and GGG conditions with 2(experience: singer and nonsinger) $\times$2(pitch shift: up and down) $\times$5(block) $\times$2(phase: early and late) as factors. The data obtained during testing revealed a significant main effect of experience and pitch shift condition, $F(1,55)=15.11, p<0.05$ and $F(1,55)=57.72, p<0.05$, respectively. Nonsingers' median F0 values were determined to be significantly lower than singers' median F0 values. The median F0 values during the shift down condition were found to be significantly higher than the median F0 values obtained during the shift up condition. A two-way interaction between pitch shift and block was significant, $F(4,220)=45.58, p<0.05$. *Post hoc* testing indicated that the median F0 values during the first block of shift trials on the shift up condition were significantly higher than all remaining shift blocks ($p<0.05$). Moreover, the median F0 values during the first block of shift trials during the shift down condition were significantly lower than shift blocks 3, 4, and 5 ($p<0.05$). Also, the median F0 values for all shift up blocks were significantly lower than the median F0 values for all shift down blocks ($p<0.05$).

Additionally, there was a significant interaction between pitch shift and phase, $F(1,55)=7.39, p<0.05$. *Post hoc* testing revealed that there were no differences between the early and late phases of the shift up and down conditions, respectively ($p>0.05$). However, the median F0 values of the early and late phases of the shift down condition were significantly higher than the median F0 values of both the early and late phases of the shift up condition ($p<0.05$). Finally, there was a significant three-way interaction between experience, pitch shift, and block, $F(4,220)=3.01, p<0.05$. *Post hoc* analysis revealed that nonsingers' first block of median F0 values during the initial block of the shift down condition was significantly lower than the median F0 values during blocks 3, 4, and 5 ($p<0.05$). Also, during the shift up condition, nonsingers' median F0 values during the initial shift block were significantly higher than the median F0 values on blocks 3, 4, and 5 ($p<0.05$). Singers' initial shift block median F0 values for both the shift down and up conditions were significantly higher and lower, respectively, than the median F0 values on shift blocks 3, 4, and 5 ($p<0.05$). Moreover, singers' median F0 values on shift down trials were significantly higher than all nonsingers' median F0 values on shift down trials ($p<0.05$). Singers' median F0 values on shift up trials were also found to be significantly higher than all nonsingers' median F0 values on shift up trials ($p<0.05$). No other significant main effects or interactions were observed.

Multiple *t*-tests (uncorrected) were also performed on the median values of the first 50 ms of each utterance. Similar to the multiple *t*-tests reported previously, three consecutive significant differences were required prior to establishing when adaptation occurred. The last five base line median F0 values were averaged to establish a base line for comparison. This value was then compared to all shifted and test trials. During the shifted up condition, singers' median F0 values were found to be different from base line at shift trial 17 (34 cents), $t(60)=-2.59, p<0.05$. In total, singers' me-

dian F0 values were different from base line on 33/50 shifted up trials. During the shifted down condition, singers exhibited adaptation responses at shift trial 13 (26 cents), $t(60)=3.94, p<0.05$. During the shifted down condition, singers' median F0 values were different than base line on 35/50 trials. Interestingly, nonsingers' median F0 values failed to significantly differ on three consecutive occasions during the shifted up or down conditions. However, nonsingers' median F0 values did differ on two consecutive trials during both the shifted up and down conditions (but only once during each). Nonsingers' median F0 values differed at shift trial 38 (76 cents) and 39 (78 cents), $t(60)=-2.73$ and 2.65, $p<0.05$, during shifted up and down conditions, respectively. Furthermore, nonsingers' median F0 values only varied from base line on 12/50 and 4/50 during the shifted up and down conditions, respectively.

Lastly, a MANOVA was performed on the median (50 ms F0) minus the mean (1500 ms) F0 values obtained during the shifted up and down AGA, FGF, and GGG conditions with 2(experience:singer and nonsinger)×2(pitchshift:up and down)×5(block)×2(phase:early and late) as factors. When the median F0 values for the first 50 ms of the shifted down and up conditions were subtracted from the mean F0 values for each respective condition, results indicated that there was a significant main effect of experience, $F(1,55)=12.52, p<0.05$. Nonsingers' F0 values were found to be significantly lower than the F0 values for singers. A main effect of pitch shift was also found to be significant, $F(1,55)=6.90, p<0.05$. F0 values during the pitch shift down condition were found to be significantly higher than the F0 values during the pitch shift up condition. Results identified a significant two-way interaction between pitch shift and block, $F(4,220)=6.86, p<0.05$. *Post hoc* testing indicated that the only difference during the pitch shift down condition was that the initial block of F0 values was significantly lower than the last block of shift trials ($p<0.05$). Moreover, the initial shift block F0 values during the shift down condition were significantly different from the F0 values of shift blocks 4 and 5 of the shift up condition ($p<0.05$). The initial block of shift trials during the shift up condition was determined to be significantly higher than shift blocks 4 and 5 ($p<0.05$), as well as the last block of trials during the shift down condition ($p<0.05$).

Finally, there was a significant three-way interaction between experience, pitch shift, and block, $F(4,220)=2.72, p<0.05$. *Post hoc* testing revealed that nonsingers' first block of F0 values during the shift down condition was significantly lower than the F0 values during the last block of shift trials ($p<0.05$). Nonsingers' F0 values during the initial block of shift up trials were significantly higher than the F0 values of blocks 4 and 5 ($p<0.05$). Singers' F0 values during the initial block of shift down trials were not significantly different from any other block of shift down trials ($p>0.05$). The F0 values of the initial block of shift up trials were also not different from any other block of shift up trials ($p>0.05$). Furthermore, nonsingers' F0 values during the shift up and down conditions were all significantly lower than singers' F0 values ($p<0.05$). No other significant main effects or interactions were observed.

IV. DISCUSSION

The purpose of this study was to investigate the sensitivity of the mechanisms that modify internal representations in singers and nonsingers when presented with subtle changes in auditory feedback while singing. In accord with our predictions, singers and nonsingers use auditory feedback to compensate for subtle manipulations in auditory feedback while singing. That is, both singers and nonsingers compensated for the altered feedback by increasing or decreasing their F0 to downward and upward shifts in feedback. The main difference between the groups was that singers are immediately and consistently more accurate when they match the target notes. Nonsingers' F0 values were consistently lower than the pitch target, regardless of the direction of the manipulation. An important observation was that singers initiated compensatory responses to altered feedback on the third shift (6 cents) trial during both upward and downward altered conditions. On the other hand, nonsingers initiated compensatory responding at shift trials 13 (26 cents) and 11 (22 cents) during upward and downward manipulations, respectively.

Furthermore, the authors examined the first 50 ms of each vocal production in order to determine how participants initiate F0 responses while singing. If participants adapt to the altered feedback by altering an internal representation that regulates their initial F0 production, their initial F0 production should be close to the F0 values produced while compensating during the previous trial. Results showed that singers, and to a lesser degree nonsingers, compensated for gradual FAF manipulations by starting subsequent utterances at similar F0 values obtained on the previous FAF trial. Thus, singers were continually updating their internal model to account for the consistently increasing or decreasing changes in F0 so they could initiate voice F0 at the desired target frequency. On the other hand, nonsingers appeared to search for the target note by starting below the auditory target of 0 cents and by increasing their F0 until they matched the note. Moreover, multiple *t*-tests failed to provide evidence to suggest that sensorimotor adaptation occurred in nonsingers during the altered feedback conditions. In the case of singers, multiple *t*-tests on the median 50 ms F0 data revealed that sensorimotor adaptation occurred at 34 and 26 cents during shifted up and down conditions, respectively.

Previous research using the FAF paradigm has typically examined vowel phonation (Burnett *et al.*, 1997; Burnett *et al.*, 1998; Burnett and Larson, 2002; Elman, 1981; Jones and Munhall, 2000, 2002, 2005; Natke *et al.*, 2003; Toyomura *et al.*, 2007), with an emphasis on compensatory responding. Not surprisingly, the results are consistent with the data collected on singing (Burnett *et al.*, 1997; Jones and Keough, 2008; Natke *et al.*, 2003, Zarate and Zatorre, 2005, 2008) in that participants compensate for manipulations in auditory feedback by increasing or decreasing their F0 in the opposite direction of the perturbation. The data obtained from the current experiment are in accord with this observation, as both singers and nonsingers altered their F0 to similar degrees while receiving altered feedback.

It has been argued that the aftereffects observed after exposure to FAF are a result of modifying an acoustic-motor representation (Jones and Munhall, 2000, 2002; Jones and Keough, 2008; Natke *et al.*, 2003). However, sensorimotor adaptation has not been extensively examined in singers. When Jones and Keough (2008) returned trained singers' auditory feedback to normal, following FAF trials, their F0 values never returned to base line levels. Thus, it appears that brief exposure to altered feedback resulted in a remapping of an internal model (Jones and Keough, 2008). In the current study, the authors also found a global remapping to subtle changes in auditory feedback. Uncorrected multiple *t*-tests revealed that aftereffects existed for both singers and nonsingers. That is, when auditory feedback was suddenly returned to normal, following exposure to FAF, singers' and nonsingers' mean test trial F0 values were significantly higher and lower (following shifted down and up feedback, respectively) than base line.

These aftereffects also generalized to a note other than the one they produced during FAF trials. For instance, singers' mean F0 values for the target note A4 during shift up and down conditions and F4 following shift down trials were significantly different from the average of the last five base line trials. Nonsingers' test F0 values were only different from base line for the target note A4 during the shift up feedback condition. Thus, participants modified an internal representation that regulates F0 control. As a consequence, the pattern of responding observed in the current study is similar to the pattern of responding that has been found in previous FAF studies (Jones and Munhall, 2000, 2002; Jones and Keough, 2008). However, when the alpha level was corrected for the multiple *t*-tests, only the aftereffects observed for singers and the generalization for the target note A4 during the shifted up condition remained significant.

Previous studies have identified aftereffects following FAF but have only examined adaptation after a period of training at the end of an experiment (Jones and Munhall, 2000, 2002; Jones and Keough, 2008). For this study, the authors tracked F0 values early in each utterance during the training period to determine whether singers and nonsingers continuously adjust their internal motor representation for F0 control in response to changes in feedback. Singers were found to match the target, on average, more accurately than nonsingers. That is, they initiated their vocal productions near the F0 frequencies that were required to compensate for the FAF experienced in previous trials. This sensorimotor adaptation in singers occurred when the discrepancy between expected F0 output and auditory feedback was approximately 30 cents. Thus, subtle discrepancies in feedback can be accounted for by the mechanisms that support singers' internal model for vocal control. On the other hand, nonsingers consistently started their productions below the target. Their F0 values began consistently below the pitch target and were adjusted (upward) until they reached the note, as best as they could. This difference between singers and nonsingers is consistent with the data obtained in Jones and Keough (2008).

One difference found in previous work investigating sustained vowel phonation and singing using FAF was the level of compensation observed. Typically, it has been reported that compensatory responses do not exceed 65 cents for shifts up to 600 cents. However, the bulk of previous research has focused on sustained vowel production while receiving random pitch perturbations (e.g., Burnett *et al.*, 1997; Donath *et al.*, 2002; Larson, 1998; Larson *et al.*, 2000; Liu and Larson, 2007). Jones and Keough (2008) reported that nonsingers fully compensated (100 cents) for highly predictable shifts in FAF almost immediately, and although singers initially showed partial compensation (approximately 65 cents) they eventually altered their F0 values to accommodate for the altered feedback. Data from the current study revealed that singers and nonsingers exhibited near perfect levels of compensation during predictable shifted up and down conditions. Thus, it appears that auditory feedback may be used in a task-dependent manner such that when achieving a particular pitch target is important, as in singing, auditory feedback guides production.

Although singers and nonsingers eventually compensated for the FAF to the same degree, the point at which each group altered their productions based on the auditory feedback differed. Singers began compensating when they detect feedback errors as large as 6 cents (1.36 Hz) during both the shift up and down conditions. This value is consistent with Sundberg's (1987) finding that trained singers can correct for production errors with an accuracy of less than 1 Hz from an intended pitch target (A4, 440 Hz). On the other hand, nonsingers initiated compensatory responses at approximately 24 cents (5.47 Hz). Data from both groups fall within the just-noticeable difference range found by Pape and Mooshammer (2006) of F0 contours for natural stimuli. Moreover, nonsingers' values are also very similar to the average threshold reported by Hafke (2008), who found that pitch shift changes were not reliably identified when they were below 26 cents. However, Loui *et al.* (2008) found that control participants' psychophysical thresholds of perception and production were around 2.0 and 2.5 Hz, respectively. Although the authors never examined the perceptual aspect directly, the data they obtained for singers and nonsingers are relatively consistent with the production threshold findings of Loui *et al.* (2008).

Based on the current findings, the authors believe that singers, due to their extensive training and experience, are more capable of compensating for subtle manipulations of auditory feedback earlier than nonsingers. Moreover, singers more readily alter their internal representations to prevent the occurrence of these errors in subsequent utterances. On the other hand, when singers detect large incongruities between perception and production they rely more on their internal model to produce the target (Jones and Keough, 2008). Singers may have developed different internal models during vocal training that provides them with a greater capacity and flexibility to control voice F0 during different task demands. The F0 control system may deem certain feedback discrepancies to be too large to be internally generated. Zarate and Zatorre (2005, 2008) found similar results when singers and nonsingers were exposed to FAF. Zarate and Zatorre (2005, 2008) asked participants to ignore the feedback and continually reproduce the target as accurately as possible. The pat-

tern of behavioral results suggests that singers could successfully ignore their altered (200 cents) feedback and continue to produce the target notes accurately. Interestingly, nonsingers were unable to ignore the feedback as their F0 responses indicated partial compensation. These findings suggest that singers' internal models are flexible in nature; the models can be adjusted to compensate for perturbations as small as 6 cents (1.36 Hz) when instructed to match the target note, or the models can be adjusted to ignore pitch errors (e.g., 200 cents) when instructed to do so by relying on the existing motor representation (Zarate and Zatorre, 2008).

Given that behavioral differences exist between singers and nonsingers during FAF, differences in neural activity between the two groups may be expected. Zarate and Zatorre (2005, 2008) found that despite differences in vocal production accuracy during normal feedback conditions, both singers and nonsingers exhibited similar functional networks for singing. These areas included the bilateral auditory cortices, bilateral primary motor cortices, the supplementary motor area (SMA), the anterior cingulate cortex (ACC), thalamus, insula, and cerebellum. These results are consistent with a previous study carried out in the Zatorre laboratory (Perry *et al.*, 1999) and by Toyomura *et al.* (2007). However, when exposed to FAF and asked to ignore the FAF, singers showed enhanced activation in the inferior parietal lobule (IPL), superior temporal gyrus (STG), superior temporal sulcus (STS), and right insula (Zarate and Zatorre, 2005). On the other hand, enhanced activation in the ACC, STS, insula, putamen, pre-SMA, and IPL was observed in singers when they were directly asked to compensate for the FAF (Zarate and Zatorre, 2005). The authors conclude that the additional activation of the STG and the STS in singers during FAF conditions is suggestive of an increased perceptual analysis of the incoming signal (Zarate and Zatorre, 2005). In addition, the authors point out that an increased activation in the putamen of singers during both ignore and compensate conditions suggests that singers are relying on well-defined internal representations to sing the targets while receiving FAF (Zarate and Zatorre, 2008). Indeed, it may be the case that singers rely on more than one internal model to regulate voice F0.

V. Conclusion

Overall, results showed that singers and nonsingers compensated for FAF to a similar degree. On the other hand, singers were more accurate in their pitch productions, as their F0 values were consistently closer to the intended pitch target. Singers compensated for FAF when auditory feedback was manipulated up or down by 6 cents, whereas nonsingers compensated when feedback was shifted upward by 26 cents and downward by 22 cents. Additionally, examining the first 50 ms of vocal productions during FAF trials showed that singers continued to initiate F0 values near the intended target frequency. The authors argue that this resulted from the online recalibration of an internal model regulating voice F0. Nonsingers were found to consistently initiate F0 productions below the pitch target and to increase their F0 until they matched the note. Thus, singers and nonsingers rely on an internal model to regulate voice F0, but singers' models appear to be more sensitive in response to subtle discrepancies in auditory feedback.

It appears that using absolute targets in a predictable FAF paradigm permitted a unique examination of F0 control. For instance, examining sensorimotor adaptation during singing has proven to be informative in testing the acoustic-motor mapping of F0 control in musically untrained and trained individuals. Using a specialized subgroup of the population may allow us to better understand how sensitive internal models are to FAF and how these motor commands are represented neurologically. Also, it is arguable that voice F0, while singing, is represented by multiple internal models that correspond to specific musical notes. Indeed, current research in our laboratory examines how sensory feedback can be used to continuously recalibrate multiple internal models regulating F0 while singing.

ACKNOWLEDGMENTS

The National Institute of Deafness and Communicative Disorders Grant No. DC-08092 and a grant from the Natural Sciences and Engineering Research Council of Canada supported this research.

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.

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

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**, 3153–3156.

Burnett, T. A., and Larson, C. R. (**2002**). "Early pitch-shift response is active in both steady and dynamic voice pitch control," J. Acoust. Soc. Am. **112**, 1058–1063.

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.

Donath, T. M., Natke, U., and Kalveram, K. T. (**2002**). "Effects of frequency-shifted auditory feedback on voice *F*0 contours in syllables," J. Acoust. Soc. Am. **111**, 357–366.

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.

Finney, S., and Palmer, C. (**2003**). "Auditory feedback and memory for music performance: Sound evidence for an encoding effect," Mem. Cognit. **31**, 51–64.

Hafke, H. Z. (**2008**). "Nonconscious control of fundamental voice frequency," J. Acoust. Soc. Am. **123**, 273–278.

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.

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

Jones, J. A., and Keough, D. (**2008**). "Auditory-motor mapping for pitch control in singers and nonsingers," Exp. Brain Res. **190**, 279–287.

Jones, J. A., and Munhall, K. G. (**2000**). "Perceptual calibration of F0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**, 1246–1251.

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.

Jones, J. A., and Munhall, K. G. (**2005**). "Remapping auditory-motor representations in voice production," Curr. Biol. **15**, 1768–1772.

Kawahara, H. (**1998**). In *Computational Auditory Scene Analysis*, edited by D. F. Rosenthal and H. G. Okuno (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 335–349.
Kawahara, H., Masuda-Katsuse, I., and de Cheveigne, A. (**1999**). "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," Speech Commun. **27**, 187–207.
Kirchner, J. A., and Wyke, B. D. (**1965**). "Articular reflex mechanisms in the larynx," Ann. Otol. Rhinol. Laryngol. **74**, 749–768.
Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.
Larson, C. R. (**1998**). "Cross-modality influences in speech motor control: The use of pitch shifting for the study of f0 control," J. Commun. Disord. **31**, 489–503.
Larson, C. R., Altman, K. W., Liu, H., and Hain, T. C. (**2008**). "Interactions between auditory and somatosensory feedback for voice F0 control," Exp. Brain Res. **187**, 613–621.
Larson, C. R., Burnett, T. A., Kiran, S., and Hain, T. C. (**2000**). "Effects of pitch-shift velocity on voice F0 responses," J. Acoust. Soc. Am. **107**, 559–564.
Lee, B. S. (**1950**). "Effects of delayed speech feedback," J. Acoust. Soc. Am. **22**, 824–826.
Liu, H., and Larson, C. R. (**2007**). "Effects of perturbation magnitude and voice F level on the pitch-shift reflex," J. Acoust. Soc. Am. **122**, 3671–3677.
Loui, P., Guenther, F. H., Mathys, C., and Schlaug, G. (**2008**). "Action-perception mismatch in tone-deafness," Curr. Biol. **18**, R331–R332.
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.
Oller, D. K., and Eilers, R. E. (**1988**). "The role of audition in infant babbling," Child Dev. **59**, 441–449.
Pape, D., and Mooshammer, C. (**2006**). "Is intrinsic pitch language dependent? Evidence from a cross-linguistic vowel pitch experiment (with additional screening of the listeners' DL for music and speech)," in MULTILING-2006, Paper No. 018.
Perry, D. W., Zatorre, R. J., Petrides, M., Alivisatos, B., Meyer, E., and Evans, A. (**1999**). "Localization of cerebral activity during simple singing," NeuroReport **10**, 3979–3984.
Proteau, L., Marteniuk, R. G., Girouard, Y., and Dugas, C. (**1987**). "On the type of information used to control and learn an aiming movement after moderate and extensive training," Hum. Mov. Sci. **6**, 181–199.
Proteau, L., Marteniuk, R. G., and Levesque, L. (**1992**). "A sensorimotor basis for motor learning: Evidence indicating specificity of practice," Q. J. Exp. Psychol. **44A**, 557–575.
Purcell, D. W., and Munhall, K. G. (**2006**). "Compensation following real-time manipulation of formants in isolated vowels," J. Acoust. Soc. Am. **119**, 2288–2297.
Sapir, S., McClean, M. D., and Larson, C. R. (**1983**). "Human laryngeal responses to auditory stimulation," J. Acoust. Soc. Am. **73**, 315–321.
Sundberg, J. (**1987**). *The Science of the Singing Voice* (Northern Illinois University Press, Dekalb, IL).
Toyomura, A., Koyama, S., Miyamaoto, T., Terao, A., Omori, T., Murohashi, H., and Kuriki, S. (**2007**). "Neural correlates of auditory feedback control in human," Neuroscience **146**, 499–503.
Watts, C., Murphy, J., and Barnes-Burroughs, K. (**2003**). "Pitch-matching accuracy of trained singers, untrained subjects with talented singing voices, and untrained subjects with non-talented singing voices in conditions of varying feedback," J. Voice **17**, 187–196.
Wyke, B. D. (**1974**). "Laryngeal neuromuscular control systems in singing: A review of current concepts," Folia Phoniatr (Basel) **26**, 295–306.
Yates, A. J. (**1963**). "Delayed auditory feedback," Psychol. Bull. **60**, 213–251.
Yoshida, Y., Saito, T., Hirano, M., and Kanaseki, T. (**1989**). "Laryngeal sensory innervation: Origins of sensory nerve fibers in the nodose ganglion of the cat," J. Voice **3**, 314–320.
Zarate, J. M., and Zatorre, R. J. (**2005**). "Neural substrates governing audio-vocal integration for vocal pitch regulation in singing," Ann. N.Y. Acad. Sci. **1060**, 404–408.
Zarate, J. M., and Zatorre, R. J. (**2008**). "Experience-dependent neural substrates involved in vocal pitch regulation during singing," Neuroimage **40**, 1871–1887.

Perceptual contributions of the consonant-vowel boundary to sentence intelligibility[a)]

Daniel Fogerty[b)] and Diane Kewley-Port
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

(Received 22 September 2008; revised 3 June 2009; accepted 4 June 2009)

Although research has focused on the perceptual contribution of consonants to spoken syllable or word intelligibility, in sentences vowels have a distinct perceptual advantage over consonants in determining intelligibility [Kewley-Port *et al.*, J. Acoust. Soc. Am. **122**, 2365–2375 (2007)]. The current study used a noise replacement paradigm to investigate how perceptual contributions of consonants and vowels are mediated by transitional information at segmental boundaries. The speech signal preserved between replacements is defined as a glimpse window. In the first experiment, glimpse windows contained proportional amounts of transitional boundary information that was either added to consonants or deleted from vowels. Results replicated a two-to-one vowel advantage for intelligibility at the traditional consonant-vowel boundary and suggest that vowel contributions remain robust against proportional deletions of the signal. The second experiment examined the combined effect of random glimpse windows not locked to segments and the distributions of durations measured from the consonant versus vowel glimpses observed in Experiment 1. Results demonstrated that, for random glimpses, the cumulative sentence duration glimpsed was an excellent predictor of performance. Comparisons across experiments confirmed that higher proportions of vowel information within glimpses yielded the highest sentence intelligibility. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3159302]

PACS number(s): 43.71.Es, 43.71.Gv [JES] Pages: 847–857

I. INTRODUCTION

All languages share a common feature: the use of both consonant and vowel sounds (Ladefoged, 2001). Intelligibility research has focused on these two fundamental groups of speech sounds, attempting to tease apart the relative perceptual contributions of each. The traditional view of these contributions has been stated clearly, as "it has become almost a commonplace statement in intelligibility testing that most of the information in speech is carried by the consonant sounds" (Owens *et al.*, 1968, p. 648). More recently, this traditional view has been questioned by Cole *et al.* (1996) and Kewley-Port *et al.* (2007). Both studies demonstrated an overwhelming advantage of vowels compared to consonants in sentence intelligibility testing using a segment replacement paradigm, also described as "noise replacement." This paradigm turns successive portions of the speech signal on and off, effectively alternating the speech signal with either silence or noise. The "on" portions of speech are defined as glimpses (Cooke, 2003). Noise replacement has been used previously to examine the specific perceptual contributions of certain speech acoustics contained in consonant or vowel segments (e.g., Kewley-Port *et al.*, 2007), as well as, the perception of interrupted speech (e.g., Miller and Licklider, 1950). While these studies were able to isolate the perceptual contributions of acoustic events present during the segment of interest, it is well known that acoustic cues for segments are distributed across consonant-vowel (C-V) boundaries (as shown for stop consonants, Liberman *et al.*, 1967). It is also clear from work with silent-center syllables that the dynamic portions of the speech signal at these C-V boundaries provide important contributions (Strange, 1987; Strange *et al.*, 1983).

In the current study, Experiment 1 sought to replicate and extend past findings by investigating how the perceptual contributions of consonants and vowels are affected by systematically varying amounts of acoustic information at the C-V boundary. Experiment 2 was conducted as a control to isolate potential confounds of the noise replacement paradigm, in order to confirm that the perceptual pattern revealed in Experiment 1 was a result of spectro-temporal information in segments, and not of particular methodological properties. In addition, Experiment 2 compared the perception of randomly ordered glimpses, the preserved portions of speech between replacements, to glimpses locked to specific segments in Experiment 1.

Segmentation of the highly dynamic acoustic waveform into discrete units corresponding to consonants and vowels is inherently ambiguous. Particularly problematic is the assignment of formant transitions to either the consonant or vowel segment, as these transitions provide information about both consonants and vowels (Liberman *et al.*, 1967). Normal coarticulation means that acoustic cues for consonants and vowels largely overlap; the perceptual effects of which may be observed over multiple segments (see Kent and Minifie, 1977). Therefore, it may be that a discrete division between consonants and vowels is largely a convenience (see Lade-

[a)]Portions of the data were presented at the 154th Meeting of the Acoustical Society of America [J. Acoust. Soc. Am. **122**, 2972 (2007)] and at the 156th Meeting of the Acoustical Society of America [J. Acoust. Soc. Am. **124**, 2439 (2007)].

[b)]Author to whom correspondence should be addressed. Electronic mail: dfogerty@indiana.edu

foged, 2001). Stevens and his colleagues (summarized by Stevens, 2002) long advocated that prominent acoustic landmarks may be used to effectively mark different segments within the ambiguous speech signal. Therefore, identifying the C-V boundary is a convenient methodological tool for dividing the acoustic signal into mostly vocalic and mostly consonantal portions. However, the perceptual cues present within the acoustic signals associated with consonants and vowels interact with each other. Cooper *et al.* (1952) were among the first to explore the relation between perception and acoustic cues in stop consonants, such as the frequency release of the burst and the direction of the second formant transition. They described the perception of these specific cues as dependent on their positional relationship to the neighboring vowel. Furthermore, the perception of speech sounds is highly dependent on the acoustic-phonetic context (Miller, 1994). Therefore, the perception of consonants and vowels and their contributions to sentence intelligibility appear to be the result of distributed and overlapping cues.

The current study uses the fundamental grouping of speech sounds into consonants and vowels to globally explore how the acoustic information contained within these segments contributes to sentence intelligibility. In addition, as the perceptual information of these segments is highly distributed and overlapping, this study investigates how segmental contributions change with the addition or deletion of transitional information between them. The current study was not designed to redefine discrete segmental boundaries but rather to investigate the relative perceptual contributions of these two primary categories of speech sounds.

A. Previous research on the relative importance of consonants and vowels

There have been several approaches to understanding the relative contributions of consonants and vowels to speech perception. Studies that have investigated the consonant-to-vowel ratio (i.e., the difference between consonant and vowel intensity levels) have found that amplifying the intensity of the consonant relative to the vowel in CVCs enhances recognition of voiced stops (Freyman and Nerbonne, 1989; Freyman *et al.*, 1991), which corroborates well with the importance of consonant acoustics for speech intelligibility. However, Sammeth *et al.* (1999) used a different approach that controlled for the intensity of three stop consonants. They maintained the intensity level of the consonant in CV syllables while decreasing the intensity level of the vowel by 6 or 12 dB. Bringing the consonant-to-vowel ratio to 0 dB did not improve identification performance of the consonants. However, in some conditions the consonants were effectively presented in isolation, representing a maximum attenuation for the vowel. Reintroduction of the vowel produced a large and significant improvement in identification, suggesting that the presence of the vowel was important for the successful identification of the consonants.

Much of the work that has championed the importance of consonant acoustics over that of the vowel has investigated this issue at the level of the syllable structure. Owren and Cardillo (2006) took this one step further by investigating multisyllabic words in isolation. They used a segment replacement paradigm to present consonant-only words or vowel-only words, where segments and formant transitions were replaced by silence. Listeners heard pairs of words and were asked to make two judgments: whether the two words shared the same meaning and whether they were spoken by the same talker. Results demonstrated that consonant-only words yielded higher d' scores for judging the meaning; vowel-only words had higher scores for judging the talker. Owren and Cardillo (2006) concluded that consonants may be more important for intelligibility, while vowels carry more indexical information about the talker.

Research at the sentence level has revealed a much different picture. In a preliminary report, Cole *et al.* (1996) used a segment replacement paradigm where segments were either replaced by speech-shaped noise, a harmonic complex, or silence. Words repeated correctly were scored. Their results suggested that vowel-only sentences maintained a two-to-one advantage over consonant-only sentences, regardless of the type of segmental replacement. Furthermore, this advantage remained, even after 10 ms was deleted from the onset and offset of the vowels. This two-to-one advantage of vowels was later replicated by Kewley-Port *et al.* (2007) when normal-hearing listeners were presented sentences at 70 dB sound pressure level (SPL), as well as for elderly hearing-impaired listeners at a higher presentation level of 95 dB SPL. For young normal-hearing listeners, the 95 dB SPL level resulted in an improvement for recognition of consonant-only sentences, but participants still maintained significantly better performance for vowel-only sentences. This advantage for vowels is even more remarkable considering that vowels actually comprise at least 10% less of the overall sentence duration than consonants (Ramus *et al.*, 1999).

As perceptual cues are highly distributed throughout the speech signal, it could be that certain subsegmental information plays an important role as well. Lee and Kewley-Port (2009) investigated four types of subsegmental information that presented equal proportions of consonants and vowels in each condition. Segmental onset and offset, as well as transitions and center portions, were investigated. These portions of segmental information were based relative to traditional segmental boundaries. Performance based on words repeated correctly was equal across all conditions, indicating no advantage for any type of subsegmental information when equal proportions of consonant and vowel information were maintained. However, while transitional and center portions of segments may make equal perceptual contributions to overall sentence intelligibility, it remains to be seen how these two different types of acoustic signals interact.

The purpose of the current study was to investigate how transitional information at the C-V boundary as traditionally defined facilitates the perceptual contributions made by consonants and vowels to sentence intelligibility. It has become widely accepted that transitional dynamics at the C-V boundary contains important information for perception (Liberman *et al.*, 1967; Strange, 1987; Strange *et al.*, 1983), but it is less clear how this information relates to the perceptual contributions of consonants and vowels. Our investigation explores

how certain types of speech information contained in consonants, vowel centers, and the transitions between them interact in sentential contexts.

B. Glimpsing speech during noise replacement

In studies that use noise replacement to isolate perceptual contributions, it may be that preserved perceptual cues in the signal are not the only performance factor, but also how those cues interact with a listener's ability to integrate information across preserved portions of the speech signal. The current study examined these perceptual consequences that are inherent in interrupted speech produced during noise replacement. The portions preserved between replacements are described here as glimpses, taken from the literature on interrupted speech in noise. Cooke (2003) defined a glimpse as "an arbitrary time-frequency region which contains a reasonably undistorted view of the target signal" (p. 579). In everyday listening conditions, it may be most beneficial for a listener to focus on the glimpses of speech between background noise fluctuations, thereby attending to the preserved speech information rather than trying to extract cues from a noisy signal. The speech intelligibility index has recently been modified to include consideration of these temporal fluctuations and now provides a good account for speech recognition thresholds in interrupted speech (Rhebergen and Versfeld, 2005). The redundancy of the speech signal makes the use of such glimpses effective in speech perception (Cooke, 2006).

In glimpse studies, in order to maintain the same amount of speech information presented across all conditions there is an acoustic trade-off between interruption rate (i.e., the frequency of available glimpses) and glimpse window width (i.e., the duration of the available speech information). Miller and Licklider (1950) demonstrated that perceptual performance for monosyllabic words is affected nonmonotonically as the glimpse rate is increased and the glimpse window width is decreased with the best performance at a glimpse rate around 10 Hz. Therefore, properties of glimpses, both in rate and size, may effect perception in noise replacement studies. This may also be related to the type of speech material used, as the perception of speech continuity is maintained at shorter glimpse durations for discourse than for isolated words (Bashford and Warren, 1987) and is associated with increased speech intelligibility when higher-order contextual cues are available (Bashford *et al.*, 1992).

In order to investigate the segmental contributions of consonants and vowels independently of glimpse contributions, a second experiment was conducted. Experiment 2 used the same distributions of glimpse window durations measured from consonants and vowels, but randomly reordered them, thereby, creating glimpses with random portions of both consonants and vowels. This allowed for the direct comparison of (1) any difference in performance due to the different glimpse durations measured from consonants and vowels and (2) performance when windows were either locked to specific segmental information or when they contained random acoustic cues.

II. EXPERIMENT 1: SEGMENTAL CONTRIBUTIONS

This first experiment was designed to answer some questions raised in previous studies conducted in our laboratory (Kewley-Port *et al.*, 2007; Lee and Kewley-Port, 2009). Results showed that using traditional segmental rules, vowels contributed more than consonants to the perception of sentences (Kewley-Port *et al.*, 2007), and that there appeared to be no advantage of transitional versus quasi-steady-state information when the relative proportion of consonant-to-vowel segment durations remained constant (Lee and Kewley-Port, 2009). However, given that perceptual information is distributed across the transitional area at the C-V boundary, a major unanswered question concerns the dominance of vowel information obtained in previous studies in relation to the allocation of transitional C-V boundary information. Thus, the present investigation systematically shifted the C-V boundary to examine how transitional information contributes to the perceptual roles of consonants and vowels for sentence intelligibility.

A. Listeners

Twenty young (age 19–30, $M=24$) normal-hearing listeners (six male, fourteen female) were paid to participate in the study. All participants were native speakers of American English and had pure-tone thresholds bilaterally at no greater than 20 dB HL at octave intervals from 250 to 8000 Hz (ANSI, 1996).

B. Stimuli

1. Speech presented in sentences

The same 42 sentences used in previous studies (Kewley-Port *et al.*, 2007; Lee and Kewley-Port, 2009) were selected from the TIMIT database (Garofolo *et al.*, 1990, www.ldc.upenn.edu) to be used in the current study. These sentences were spoken by 21 male and 21 female talkers from the North Midland dialect region that matched the dialect region of daily exposure for the participants in this study (i.e., Indianapolis, Indiana and further north).

Sentences contained an average of 8 words (range 6–10 words) and 33 phonemes (11 vowels, 22 consonants). The C-V boundaries are specified by the TIMIT database. Phonemes were identified by traditional segmental boundaries, although TIMIT first employed CASPAR (Leung and Zue, 1984) automatic segmentation to identify the boundaries, which experienced phoneticians later confirmed (see Zue and Seneff, 1988). The TIMIT segmentation methods included rules such as (1) finding highly salient and abrupt acoustic changes to mark phoneme boundaries for stops and (2) dividing formant transitions in half during slow periods of change for liquids (presumably because these transitions provide information regarding both phonemes). Kewley-Port *et al.* (2007) verified the segmental boundaries provided by the TIMIT corpus and added three minor rules appropriate for identifying the vowels and consonants in sentences. These rules were used in the present study and are noted below using the TIMIT arpabet.

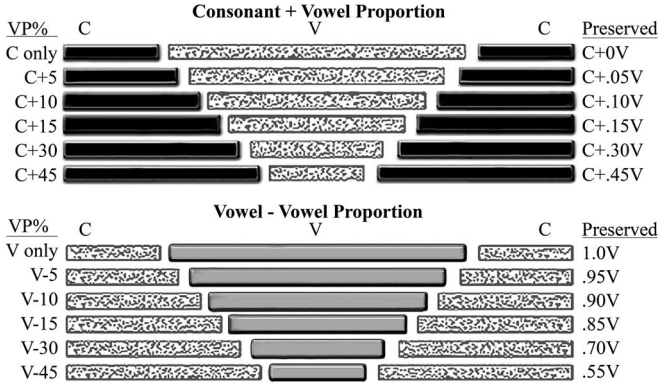


FIG. 1. Schematic of noise replacement conditions depicted for a single CVC within the sentence (not drawn to scale). The C only and V only (for C+0 and V-0 conditions) uses the original TIMIT C-V boundaries. Solid bars indicate the speech portions presented. Stippled bars indicate noise replacement.

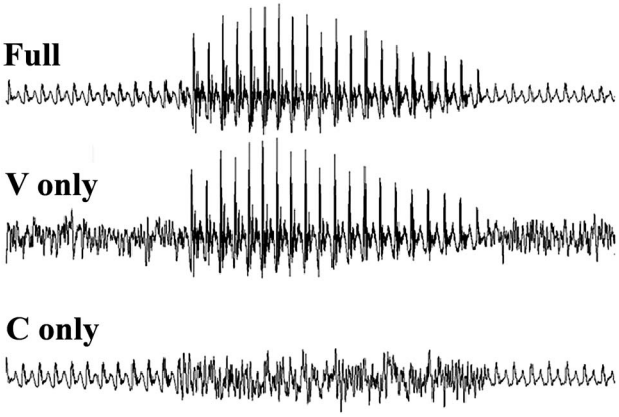


FIG. 2. Waveforms demonstrating two noise replacement conditions and the full waveform of the word "mean". V only = vowel-only; C only = consonant-only.

(1) Stop closures (e.g., "bcl") were combined with the stop (e.g., "b") and treated as a single consonant.
(2) Syllable final V+[r] as in "deer" were transcribed in TIMIT as a vowel plus the consonant [r]. These V+[r] transcriptions were considered as a single rhotocized 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 part of a vowel string.

For the purposes of the noise replacement paradigm used here and elsewhere (see Kewley-Port *et al.*, 2007), consonant strings were treated as a single unit for replacement, as were vowel strings. Therefore, the experimental manipulation of C-V boundaries was only between consonant and vowel units. The acoustic boundary was shifted by a proportion of the vowel duration (VP), thereby, reassigning vowel transitions, because previous research had demonstrated relatively low performance for consonant-only stimuli (Kewley-Port *et al.*, 2007). It was expected that manipulating by a proportion of the vowel would yield the most interpretable data by avoiding floor performance. This method allows direct investigation of how transitional information typically assigned to vowel segments influences the perceptual contributions of vowels and consonants.

Two processing strategies each created a different stimulus type. One strategy preserved the consonant information and added some proportion of the vowel transitions (C+VP), replacing the vowel centers with noise. The second strategy preserved only a proportion of the vowel center information, replacing the consonants and remaining vowel transitions at the C-V boundary with noise (V−VP). These processing strategies were manipulated by shifting the C-V boundary by six vowel proportion amounts (2×6 design), yielding a total of 12 conditions. Figure 1 shows a schematic of these 12 conditions. The schematic example shows the noise replacement for a single CVC; however, this method was applied to all consonant and vowel units across the entire sentence. The conditions are labeled in two ways on the figure, first according to how the C-V boundary was shifted (e.g., C+5, meaning the original consonant segment plus 5% of the vowel), and second, by an equation of the proportion of the segments preserved (e.g., C+.05V). The rest of this manuscript will use the first label type in referring to these conditions.

2. Noise presented in sentences

Segments were replaced by a speech-shaped noise generated in MATLAB that was based on a standard long-term-average speech spectrum (ANSI, 1969). The noise was designed to be flat from 0 to 500 Hz and had a −9 dB/octave roll-off above 500 Hz. The noise level was presented 16 dB below the average sentence level (21 dB below the vowel with loudest rms). This level was used in a previous study for consonant noise replacement (Kewley-Port *et al.*, 2007) to avoid any potential for temporal masking of consonant segments. Every segment was replaced using a unique noise. Figure 2 displays the waveform of a single CVC, "mean," excised from an experimental sentence under vowel-only and consonant-only noise replacement conditions with the full waveform as reference.

C. Calibration

Similar procedures to Kewley-Port *et al.* (2007) and Lee and Kewley-Port (2009) were administered to verify signal levels. First, scripts were written in MATLAB to verify that all 42 test sentences had similar average rms levels (i.e., within ±2 dB). Second, a MATLAB script was used to find the most intense vowel across all sentences and then iterated that vowel to produce a calibration vowel of 4 s. As these stimuli were previously used to evaluate sentence intelligibility with listeners who had hearing impairment, all sentences and the calibration vowel were filtered by a low-pass FIR filter that was flat to 4000 Hz with a 3 dB cutoff at 4400 Hz, and a 200 dB/octave slope in a Tucker-Davis Technologies programmable filter. Therefore, this allows direct comparison of this study with previous ones (Kewley-Port *et al.*, 2007; Lee and Kewley-Port, 2009). The sound level for the calibration vowel was set to 100 dB SPL through ER-3A insert earphones in a HA-2.2-cm^3 coupler using a Larson Davis model

2800 sound level meter with linear weighting. Relative to the loudest calibration vowel, the mean of the distribution of the loudest vowels in the other sentence was 95 dB SPL. All sentences and the calibration vowel were passed through a programmable attenuator to present the stimuli at 70 dB SPL, being the nominal level referenced in this study. An additional low-level background noise was continuously presented during testing. The purpose of this noise was to reduce transients between speech and the replacement noise. This background noise was generated by a Tucker-Davis Technologies waveform generator, and was also low-pass filtered at 4400 Hz. The level of this noise was presented at 50 dB less than the calibration vowel (100 dB SPL) measured using the same equipment described above.

D. Procedure

Listeners were seated in a sound attenuated booth and listened to sentences presented monaurally to the right ear via ER-3A insert earphones. Test stimuli were controlled by TDT system II hardware connected to a PC computer running a MATLAB stimulus presentation interface. Each listener was instructed regarding the task using verbal and written instructions and completed a familiarization task prior to experimental testing. The familiarization task consisted of six non-experimental sentences. The first two sentences were unprocessed. The last four processed sentences were presented in a random order of vowel-only and consonant-only conditions. The processed version was presented first followed by the full unprocessed version of the same sentence. The listener was requested to repeat aloud all of the words they heard for both versions of the sentence in order to ensure they understood the experimental task. All participants moved on to the experiment following these familiarization sentences regardless of performance.

During the experimental testing, each participant was randomly assigned to one of two condition subsets, each of which contained 6 of the 12 experimental conditions. Each participant heard 14 sentences per condition (114–115 words; ~462 phonemes). The six experimental conditions included both processing strategies (V−VP and C+VP) for three vowel proportions. Sentences were presented in two blocks with different fixed orders of three quasi-randomly assigned conditions each, resulting in participants listening to a total of 84 sentences. Past research demonstrated little benefit of hearing the sentence presented twice, even when presented sequentially (Kewley-Port *et al.*, 2007). Here, the same sentence was never presented in the same block to participants. The possibility of a sentence repetition effect is explored in the results.

The task for each subject was to repeat each sentence aloud. Listeners were encouraged to guess, without regard to whether their responses made logical sentences. No feedback was provided. The experimenter scored responses on paper during the session and digitally recorded them for later reanalysis by an independent scorer (Inter-rater agreement on 10% of responses: 98.8%). Only words repeated exactly correct were scored (e.g., no missing suffixes). For example, a response of "pool" would be judged as incorrect if the target sentence had used the plural form "pools." In addition, words were allowed to be repeated back in any order to be counted as correct repetitions.

E. Results and discussion

1. Descriptive analysis

In the first analysis, possible effects of learning were investigated because each sentence was presented twice to listeners, each time under a different condition. Five sentences that occurred within the first ten sentences presented in block 1 and within the last ten sentences in block 2 were selected for analysis. For each sentence, the average percent correct across each of the participants was calculated. This average score was transformed to a z-score for its particular segmental condition. No significant difference in performance was found, $p=0.98$; normalized scores corresponded to a performance increase of 1.6% when the sentence was presented a second time. Furthermore, when considering z-scores for all of the sentences, there was no significant difference between blocks. These results indicate that listeners' performance remained consistent across trials and were unaffected by hearing a sentence presented a second time in a different condition. Thus, reported results appear reliable due to listeners having stable, repeatable performance and were not affected by procedural learning across the two blocks.

Consonants and vowels were manipulated in different directions to investigate the contributions of transitional information. This specific experimental design required a series of six paired samples t-tests to be conducted using Bonferroni corrections for multiple comparisons to compare performance across each of the vowel proportions tested. Overall, results indicated that the vowel-only (V only) condition ($M=64.7\%$ words correct) had a significant [$t(9)=-9.7$, $p<0.001$], two-to-one advantage over the consonant-only (C only) condition ($M=30.5\%$ words correct), replicating Cole *et al.* (1996) and Kewley-Port *et al.* (2007). The other five t-tests between consonant (C+VP) and vowel (V−VP) conditions for each VP pair were significantly different at VP=5% [$t(9)=-5.6$, $p<0.001$], VP=10% [$t(9)=-5.7$, $p<0.001$], VP=30% [$t(9)=4.2$, $p<0.008$], and VP=45% [$t(9)=14.9$, $p<0.001$]. However, no significant difference was found when VP=15% [$t(9)=-1.5$, $p=0.18$], indicating that consonant (C+15) and vowel (V−15) conditions approached similar intelligibility. Figure 3 shows the data points and trends for the consonant and vowel conditions drawn with equal fit ($R^2=0.98$) using least squares regression. A linear function approximated the C+VP conditions, while a cubic function was required to provide the same fit for the V−VP conditions. Thus, C+VP conditions increased linearly with the addition of proportional vowel information. In contrast, V−VP conditions, with a nearly constant function around 55% correct, appeared to remain robust against proportional decreases in information until 30% of the vowel was removed.

These results suggest that increasing transitional information at C-V boundaries linearly predicts perceptual accuracy when consonants are presented. However, for vowels,

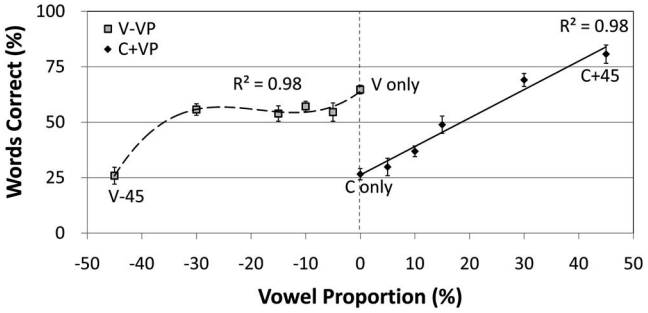


FIG. 3. Results for the 12 experimental conditions in Experiment 1 plotted according to the vowel proportion by which the C-V boundary was shifted. The original TIMIT C-V boundary is at 0% VP, marked by the dashed vertical line. Error bars display the standard error of the mean.

this transitional information at the C-V boundary appears to provide little added perceptual benefit, suggesting that the C-V boundary transitions provide information redundant with the contributions from vowel centers, when measuring sentence intelligibility. Similar results were found with CVCs by Strange *et al.* (1983) who suggested that information for the vowel is distributed across the changing acoustics of the syllable. In our study, the resilience of vowel centers remained until 30% of the transitional vowel information at the boundaries was deleted. This corresponded to when intelligibility was on average 56% correct for vowels and 69% for consonants. Note that this 30% proportional C-V shift maximized performance for both consonants and vowels.

2. Sentence level analysis

Figure 4 shows on average, how much of the total sentence was preserved across the 12 conditions. The total sentence duration on average was 2.48 s. Sentence-level measurements for the original TIMIT boundaries demonstrated that consonants comprised 55% of the total sentence duration, while vowels only comprised 45%. While individual consonants have a shorter duration than vowels, more consonants ($M=22$) occurred per sentence than vowels ($M=11$). On average, 74% of consonants occurred in consonant strings, while only 12% of vowels occurred in vowel strings. When 30% of the vowel is removed in the V−30 condition performance is at 56% correct. Performance in the consonant conditions does not surpass this 56% performance level until a 30% shift in the boundary for the (C+30), where performance reaches 69% correct. Note that for this shift, 2/3 of the sentence is presented in this consonant condition, compared to only 1/3 presented in the corresponding vowel condition just mentioned. Clearly, the cumulative amount of the sentence presented is not an underlying factor that provides the vowel conditions an advantage. When considering perceptual performance, it is notable that vowels contribute twice as much to sentence intelligibility when the original C-V boundary is used despite preserving less of the overall sentence duration (only 45% for vowels). When considering duration of the presented speech signal, consonants must present twice as much of the total sentence duration to surpass performance of the vowel.

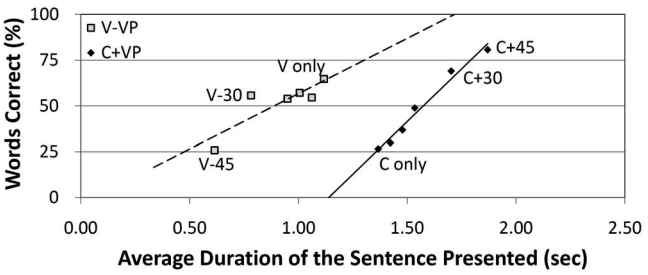


FIG. 4. The average amount of the sentence preserved across the 12 conditions with the rest of the sentence replaced by noise. The amount presented was distributed over the entire sentence according to the specific noise replacement condition. The length of the average sentence was 2.48 s.

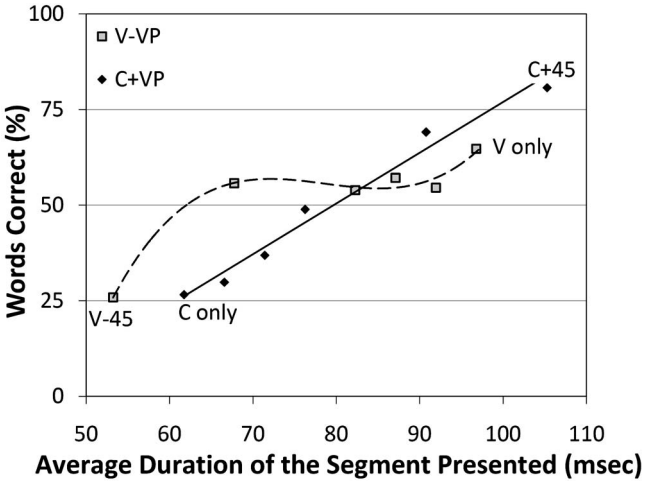


FIG. 5. The average duration of an individual segment in each condition (see text for details).

3. Segmental level analysis

Another way to analyze the results is by each individual segment, regardless of whether it occurred as part of a string of segments. This analysis corresponds to examining the effects relative to the average segment duration rather than average duration of each glimpse of the speech waveform. Figure 5 displays the percent of words correct relative to the duration of signal information present in the average speech segment calculated separately for each condition. For the V only and C only conditions, this corresponds to the average vowel or consonant duration based on the original TIMIT boundaries. On average, the C only condition provided a shorter segment duration ($M=62$ ms) than the V only condition ($M=97$ ms). However, for the rest of the conditions, a segment is defined according to the noise replacement paradigm used in this study. For example, a segment in the C+45 condition is the original consonant plus 45% of the vowel duration, with an average duration per segment of 105 ms.

When the amount of speech information was varied by a proportion of the vowel, intelligibility as a function of segment duration for the consonant and vowel conditions crossed. Specifically, at the segmental level, equal perceptual contributions to sentence intelligibility occurred when the signal duration on average for consonants and vowels was equal to 83 ms. Listener performance was matched at this

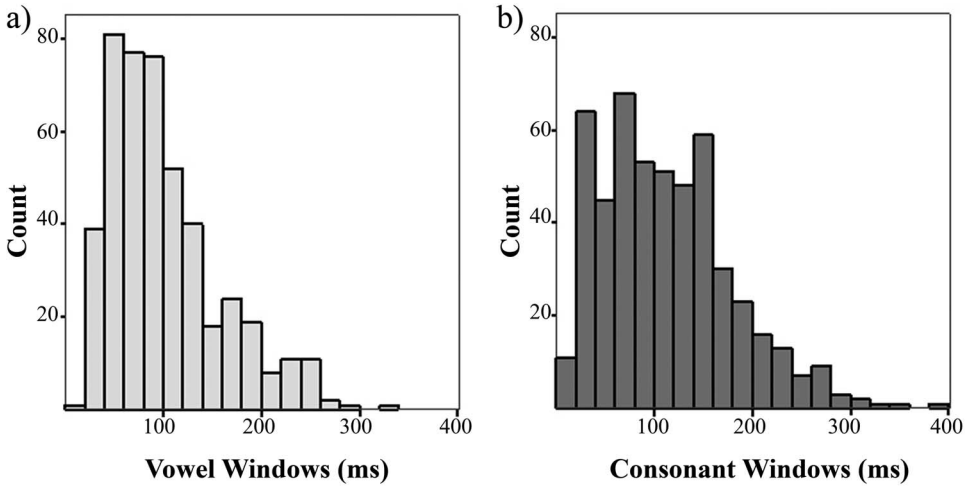


FIG. 6. Glimpse window distributions measured from all segments which occurred across all 42 sentences for (a) vowel glimpse windows and (b) consonant glimpse windows.

point for a word identification score of 55% correct. While this absolute value of 83 ms is likely dependent on factors such as speech rate, this point of perceptual equivalence occurred nearest to our 15% VP shift in both segmental conditions. As previously discussed, there was no statistical difference in the C+15 and V−15 conditions. Thus, while a 30% VP shift of the C-V boundary maximized both consonant and vowel perceptual contributions to sentence intelligibility, a 15% VP shift most successfully (out of the conditions presented) equalized such contributions.

4. Glimpse window analysis

The durations of glimpse windows locked to consonant and vowel segments were measured across all 42 sentences to determine if these durations composed significantly different distributions for the two segmental types (see Fig. 6), thereby providing an opportunity for different perceptual contributions related to glimpse duration. While there was a large overlap in the distributions for consonant ($M=113$ ms, $SD=67.6$ ms) and vowel ($M=102$ ms, $SD=56.6$ ms) glimpse windows, there was a significant difference between the distributions [$t(460)=2.642$, $p<0.01$]. Therefore, while results for Experiment 1 clearly demonstrated that glimpse windows locked to vowel information were essential for sentence intelligibility, it may be that the results were largely related to glimpse window differences rather than to segmental differences found in the acoustic content of the windows. Experiment 2 was designed to tease apart the independent contributions of segmental acoustics and glimpse windows.

III. EXPERIMENT 2: GLIMPSE CONTRIBUTIONS

The better performance with vowels in Experiment 1 may be due to cues inherent to segment glimpse window durations rather than specific segmental acoustic cues provided in the glimpse (as vowel windows overall were significantly shorter than consonant windows). This may have occurred because performance changes as a function of combined changes in glimpse rate and duration (Miller and Licklider, 1950). Specifically, glimpses of different durations, as noted between the different durations of consonants and vowels, may affect performance in addition to the actual acoustics presented. To investigate this issue, a second experiment was designed to control for specific segmental contributions by ensuring that acoustics from both consonants and vowels were equally likely to occur within a single glimpse. Experiment 2 examined: (1) the differential contributions to sentence intelligibility of glimpse window durations measured from consonants and vowels and (2) the effect of randomly placed glimpse windows that are asynchronous with segmental information. In addition, these glimpse windows were varied according to a subset of six durational proportions selected from among the 12 conditions of Experiment 1.

A. Listeners

Ten young (age 19–25, $M=22$) normal-hearing listeners (three male, seven female) were paid to participate in the study. These listeners were different from those who participated in Experiment 1. All participants were native speakers of American English and had pure-tone thresholds bilaterally at no greater than 20 dB HL at octave intervals from 250 to 8000 Hz (ANSI, 1996).

B. Stimuli

The same sentences were used as in Experiment 1 and differed only in processing strategy. The same glimpse windows were used within each sentence, preserving both the number and duration of glimpses as calculated from the TIMIT database. The order of these windows was then quasi-randomized such that glimpse boundaries no longer corresponded with segmental boundaries. The quasi-randomization followed three rules. First, the windows alternated between windows measured from consonant units and those measured from vowel units, such that no two windows measured from the same segment type occurred in adjacent intervals; thereby avoiding two glimpse windows combining into a larger one. Second, for the majority of sentences there was an unequal number of windows between the two segmental types. In order to ensure that glimpse boundaries were maximally offset at the beginning of the sentence, either the shortest or longest window measured from the segmental type with the most windows in that sentence was randomly selected to be presented first. When there was an

equal number of consonant and vowel windows, this rule did not apply and the first segment window was selected at random from the opposite segmental type presented first in the sentence (e.g., if the sentence began with a consonant, a vowel window was selected at random). Third, when there were pause windows (i.e., silent intervals) present in the sentence, these windows were presented in the randomized order closest to their original location in the sentence in an attempt to preserve sentence prosody. Without this latter rule, the segmental window that fell within a silent period in the sentence would convey no acoustic information. For Experiment 2, it was important to preserve the same amount of speech acoustics as presented in Experiment 1; otherwise poor performance could be related to reduced amounts of speech acoustics overall. This possibility was explicitly avoided by this third rule.

As with Experiment 1, two different processing strategies were used: one that used consonant glimpse window durations and one that used vowel glimpse window durations. The durations of these windows were manipulated by three of the previous six proportions of the vowel (0%, 15%, 30%) to yield a total of six conditions. This subset of conditions was selected to explore intelligibility at glimpse durations that provided relatively stable performance for vowel conditions. The glimpse windows measured in these conditions corresponded directly to those used in Experiment 1 under the identical vowel proportion condition. The only difference here was that glimpse windows were no longer time-locked to specific segmental boundaries. Calibration for this experiment used procedures identical to those in Experiment 1.

C. Procedure

Test procedures were identical to those of Experiment 1. Each participant completed a familiarization task identical to Experiment 1, but with randomized glimpse conditions. During the experimental testing, participants again completed two different blocks of 42 sentences. Each block contained three different conditions in a fixed random order. Sentences were presented once in each block under a different condition with a different randomization of glimpse windows. Participants again repeated aloud each sentence without feedback. The experimenter scored responses on paper during the session and digitally recorded them for later reanalysis by an independent scorer (Inter-rater agreement on 10% of responses: 96.5%).

D. Results and discussion

1. Glimpse window analysis

As with Experiment 1, the effect of practice resulting from procedural learning and sentence repetition was examined to determine the reliability of participant responses. Analysis of five sentences that were presented within the first ten sentences of block 1 and the last ten sentences of block 2 were compared. No significant difference in z-scores was observed ($p=0.71$). The z-scores for these two presentations corresponded to a 3.1% increase in scores for the five sentences at the end of the second block. Thus, results for this second experiment were stable under this processing strategy.

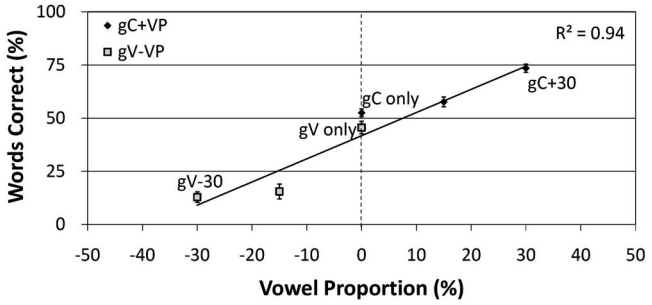


FIG. 7. Results for the six experimental conditions in Experiment 2 plotted according to the vowel proportion by which the C-V boundary was shifted. Glimpse window durations based on the original TIMIT C-V boundary are at 0% VP, marked by the dashed vertical line. Glimpse windows measured from the consonant (black diamonds) and vowel (gray squares) conditions of Experiment 1 are coded similarly to previous figures but now with the prefix "g" to indicate random glimpses. Error bars display the standard error of the mean.

Figure 7 displays the results of Experiment 2 according to how the distributions of glimpse duration were shifted by vowel proportion. All condition labels for Experiment 2 are appended with an initial "g" to indicate that they only preserve the duration of speech within a glimpse and do not contain specific segmental information. Results were analyzed by a series of Bonferroni-corrected t-tests conducted between glimpses measured from consonants and vowels at each of the three vowel proportions tested. The analysis indicated a significant difference between glimpse windows of gC+VP and gV−VP conditions at 15% [$t(9)=9.4$, $p<0.001$] and 30% [$t(9)=21.8$, $p<0.001$] VP pairs, but not at the original C-V boundary (0% VP), [$t(9)=1.7$, $p=1.88$]. Thus, the two-to-one advantage obtained with segmental windows was not observed between randomly presented windows measured from consonants versus those from vowels. These results suggest that the large difference between glimpses locked to segmental information in Experiment 1 was a result of perceptual cues contained in the segmental acoustics, not characteristics inherent in the distributions of glimpse windows. Thus, the initial conclusions of Experiment 1 are validated: the perceptual cues contained within vowel segments do contribute more to sentence intelligibility than those within consonant segments.

Figure 8 displays the six random glimpse conditions of Experiment 2 (black) plotted according to the cumulative duration of the sentence presented. As a reference, results from the corresponding conditions of Experiment 1 (gray) are also plotted. A linear trend (solid line) was fit to the six glimpse conditions tested ($R^2=0.93$). In fact, when glimpses were not locked to segmental information, the percent of the sentence presented closely predicted performance, particularly for the four longer values. That is, given the percent of the sentence presented at 45%, 55%, 62%, and 69%, the percent of the words identified was predicted by those values within 1%–5%. While such close predictions did not hold for

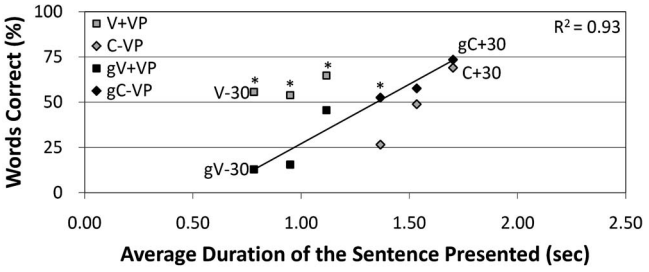


FIG. 8. The percent of words correct is displayed for the six conditions of Experiment 2 (black) and the comparison conditions of Experiment 1 (gray) as a function of the cumulative duration of the sentence preserved over the entire sentence. The black regression line is only across the six random glimpse conditions. Asterisks (*) indicate a significant difference between Experiments 1 and 2, $p<0.001$.

the two shorter durations (gV−15 and gV−30), overall, the predictability of word accuracy from only the duration of the sentence presented was remarkably high.

2. Comparison of results with Experiment 1

An analysis was conducted to examine the difference in experimental results across the two experiments. Bonferroni-adjusted t-tests between Experiments 1 and 2 for the six paired conditions (see Fig. 8) demonstrated significant differences for four conditions tested, but not for the two longest glimpse window conditions, C+15 [$t(9)=2.0$, $p=0.08$] and C+30 [$t(9)=1.3$, $p=0.24$]. Of the four conditions that were significantly different, three conditions [V−30, $t(9)=-12.8$, $p<0.001$; V−15, $t(9)=-8.5$, $p<0.001$; and V only, $t(9)=-7.8$, $p<0.001$] yielded higher performance in Experiment 1 with segmentally locked glimpses that contained only vowel information. Performance was poorer with randomly placed glimpses that only contained partial vowel information. The benefit of vowel segmental information is clearly seen in all three comparisons (V only, V−15, and V−30), but especially at the shortest glimpse duration, V−30. At the V−30 duration, providing glimpses of only vowel information in the segmental condition of Experiment 1 yields a perceptual advantage over the random glimpses of Experiment 2 by a factor of 4. The fourth significant condition, C only [$t(9)=12.6$, $p<0.001$], is a special case where segmentally locked glimpses in Experiment 1 contained no vowel information. For this condition, performance was significantly better with randomized glimpses, presumably because of the inclusion of partial vowel information. Therefore, our interpretation of the results for the four conditions is as follows: significant differences between segmental and random glimpses from the two experiments can be explained by increased amounts of preserved vowel information. Considering the similar performance of the two longest glimpse windows, these conditions (C+15 and C+30) contained the most segmental information distributed over both consonants and vowels. For these last two conditions, random glimpses apparently provided the same amount of perceptual information as is provided when larger amounts of vowel information were added to C only information in Experiment 1. Therefore, these glimpse windows are long enough for the combined acoustics of both consonants and vowels to contribute to the perception of locked or randomized glimpse windows. In summary, the presence of more vowel information in glimpses yields higher perceptual performance. Vowel information is particularly important when less of the sentence is presented.

IV. GENERAL DISCUSSION

A. Segmental contributions to sentence intelligibility

The current study investigated the perceptual contributions of vowels, consonants, and the transitions between them, using noise replacement with locked or random glimpses. Results demonstrated the importance of vowel information to sentence intelligibility as described by Cole *et al.* (1996) and Kewley-Port *et al.* (2007). Vowels as traditionally defined carry important perceptual cues that are not found in consonants. Furthermore, even truncated portions of vowels contribute strongly to sentence intelligibility despite providing much less of the overall sentence duration than consonants. Perceptual cues are distributed across the entire vowel, such that transitions into the vowel provide little additional perceptual benefit above what is provided by the vowel center alone, when the vowel center contains 70% of the total vowel. However, these dynamic transitions do provide additional perceptual information to consonants which significantly improve sentence intelligibility. Such perceptual information in the "vowel" transitions is not found in traditionally defined consonants, as indicated by the relatively low consonant-only sentence scores. Our results correspond well with the literature that has identified C-V transitions (Liberman *et al.*, 1967) and vowel context (Cooper *et al.*, 1952) as important cues for consonant recognition, and with evidence from silent-center vowels that demonstrated the importance of dynamic information for vowel recognition (Strange, 1987; Strange *et al.*, 1983).

Owren and Cardillo (2006), using a similar replacement technique with silence rather than noise, found that consonants are more important to the intelligibility of isolated words. However, it may be that vowels gain greater perceptual importance from an increased amount of contextual or stylistic factors present in sentential contexts. In sentences, speech acoustics are characterized by shorter individual segments and greater coarticulation. Thus, individual consonant segments may become less distinct, requiring more transitional dynamics for recognition. This blurring of segmental boundaries results in words, originally intelligible in sentences, becoming largely unintelligible when excised (Craig and Kim, 1990). Clearly, listeners take advantage of coarticulatory and prosodic cues in the perception of fluent sentences. In fact, some speech recognition methods explicitly define coarticulatory events between words in order to achieve higher recognition performance for continuous sentences (Giachin *et al.*, 1991).

The most sonorant segments in speech, namely the vowels, are most capable of carrying prosodic information such as pitch contours or stress. Prosody spans linguistic levels (Kent and Read, 1992), providing an opportunity for multiple perceptual and cognitive levels involved in sentence recognition to interact. For example, prosody may provide

information about syntactic structure through cues such as phrase-final lengthening. This particular prosodic cue, tied to the last stressed syllable in a major syntactic phrase, is perceptually available for listeners to use in recognizing the structure of speech (Read and Schreiber, 1982). Vowels may be particularly useful in conveying these cues for syntactic structure, as stress patterns temporally align with the onset of vowel energy (Tajima and Port, 2003). Indeed, Toro *et al.* (2008) demonstrated that participants were able to extract syntactic cues from vowels, but not from consonants. Furthermore, they support a functional dissociation of consonants and vowels, with consonants cuing the lexicon and vowels cuing syntactic structure. It may be that the syntactic cues that vowels uniquely carry facilitate top-down processes and constrain lexical activation.

Therefore, there are many types of information that vowels may carry to provide them with an advantage over consonants during the perception of compromised sentences containing only partial information. These vowel-associated cues encompass acoustic, coarticulatory cues important for consonant recognition, but may also involve perceptual cues that provide higher-order mechanisms to predict information, even though many explicit acoustic cues for segment recognition may be missing. Future studies will have to tease apart potential contributions from the complex acoustic, perceptual, and cognitive events that occur in sentences, which we as listeners navigate with relative ease in every day conversational contexts.

B. Glimpse contributions to sentence intelligibility

Even though there was a significant difference between the durational distributions of consonant and vowel glimpses, no differential performance was found for perceptual intelligibility of sentences when using these glimpse windows in a randomly ordered replacement paradigm. Instead, the amount of the speech signal presented, distributed over the length of the entire sentence, predicted perceptual performance. This result is consistent with Iyer *et al.* (2007) who used periodic glimpses to investigate the effect of noise and speech maskers. Above an 8 Hz interruption rate, the trade-off between the glimpse window duration and number of glimpses does not have much of an effect, as long as the cumulative sentence duration glimpsed (i.e., the total duration of preserved speech information) remains constant. Also consistent with the current study, Li and Loizou (2007) found that the cumulative duration of speech glimpsed predicts performance rather than glimpse window duration. By comparing performance across several glimpse conditions that preserved different frequency regions of speech, Li and Loizou (2007) also identified the frequency region of 0–3 kHz as the most important frequency region for glimpsing. This frequency region corresponds to the F1 and F2 frequency regions most important for vowel perception. Therefore, this evidence provides further support for the important contribution of vowel information to sentence intelligibility.

Experiment 2 investigated the combined perceptual effect of the distributional properties and random (i.e., aperiodic) placement of glimpse windows. In a pilot study, we separately assessed the contributions of overall sentence duration presented while avoiding differences in the distributions of consonants versus vowels. The pilot used periodic replacement based on the duty cycle that was the average of the glimpse windows (215 ms). For the consonant glimpse condition, the on phase was 113 ms, the mean glimpse window size for consonants, while the off phase was 102 ms, the mean glimpse size for vowels. For the vowel glimpse condition, the on and off phases were reversed (e.g., the on phase was the average vowel window duration of 102 ms). These on/off phases were manipulated by a proportion of the vowel as in the experiments presented here to provide a direct comparison. The pilot for all the periodic replacement conditions yielded very similar results to those obtained in Experiment 2 with aperiodic replacement, further suggesting that glimpse perception is predicted by the cumulative duration of the sentence presented. These results are consistent with Miller and Licklider (1950) who found that perception of monosyllabic words during random, aperiodic interruption was effectively the same as during periodic interruption.

Overall, there is a significant difference between the segmentally locked glimpses in Experiment 1 and the random, segmentally asynchronous glimpses used in Experiment 2. There is a significant advantage for glimpses that capture specific spectro-temporal information contained in vowel segments.

V. CONCLUSIONS

This study used a fundamental division in speech sound categories, specifically between consonants and vowels, to investigate contributions to sentence intelligibility. The present study replicated results from Cole *et al.* (1996) and Kewley-Port *et al.* (2007) that demonstrated a two-to-one advantage for sentences containing only vowels compared to sentences that contained only consonants. Furthermore, this study investigated how these contributions were mediated by transitional information into the vowel. This transitional information appears to be redundant with information at the vowel center, suggesting a similar result to Lee and Kewley-Port (2009) who found no difference in transitional versus quasi-steady-state information. However, increased transitional information did provide a linear benefit relative to consonant-only sentences, demonstrating that critical perceptual information is present in the transition. Finally, the current study demonstrated that glimpse window distributions did not differentiate performance between consonant-only and vowel-only sentences. Instead, the cumulative amount of the sentence present and the amount of vowel information preserved served as predictors for listener performance. Results confirm the importance of vowels to sentence intelligibility, as well as how dynamic transitional information at the C-V boundary significantly improves the perceptual contributions of consonants. Simply put, vowel information is essential for sentence intelligibility, particularly when less of the overall sentence duration is presented.

ACKNOWLEDGMENTS

This work was supported by NIH-NIDCD Training Grant No. T32-DC00012 and R01-02229 to Indiana University.

ANSI (**1969**). "American National Standard Methods for Calculation of the Articulation Index," ANSI S3.5-1969, American National Standards Institute, New York.

ANSI (**1996**). "Specifications for audiometers," ANSI S3.6-1996, American National Standards Institute, New York.

Bashford, J. A., Jr., and Warren, R. M. (**1987**). "Multiple phonemic restorations follow the rules for auditory induction," Percept. Psychophys. **42**, 114–121.

Bashford, J. A., Jr., Reiner, K. R., and Warren, R. M. (**1992**). "Increasing the intelligibility of speech through multiple phonemic restorations," Percept. Psychophys. **51**, 211–217.

Cole, R., Yan, Y., Mak, B., Fanty, M., and Bailey, T. (**1996**). "The contribution of consonants versus vowels to word recognition in fluent speech," in Proceedings of the ICASSP'96, pp. 853–856.

Cooke, M. (**2003**). "Glimpsing speech," J. Phonetics **31**, 579–584.

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

Cooper, F., Delattre, P., Liberman, A., Borst, J., and Gerstman, L. (**1952**). "Some experiments on the perception of synthetic speech sounds," J. Acoust. Soc. Am. **24**, 597–606.

Craig, C. H., and Kim, B. W. (**1990**). "Effects of time gating and word length on isolated word recognition performance," J. Speech Hear. Res. **33**, 808–815.

Freyman, R. L., and Nerbonne, G. P. (**1989**). "The importance of consonant-vowel intensity ratio in the intelligibility of voiceless consonants," J. Speech Hear. Res. **32**, 524–535.

Freyman, R. L., Nerbonne, G. P., and Cote, H. A. (**1991**). "Effect of consonant-vowel ratio modification on amplitude envelope cues for consonant recognition," J. Speech Hear. Res. **34**, 415–426.

Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., and Dahlgren, N. (**1990**). "DARPA TIMIT acoustic-phonetic continuous speech corpus CDROM," National Institute of Standards and Technology, NTIS Order No. PB91-505065.

Giachin, E. P., Rosenberg, A. E., and Lee, C.-H. (**1991**). "Word juncture modeling using phonological rules for HMM-based continuous speech recognition," Comput. Speech Lang. **5**, 155–168.

Iyer, N., Brungart, D., and Simpson, B. (**2007**). "Effects of periodic masker interruption on the intelligibility of interrupted speech," J. Acoust. Soc. Am. **122**, 1693–1701.

Kent, R., and Minifie, F. (**1977**). "Coarticulation in recent speech production models," J. Phonetics **5**, 115–133.

Kent, R. D., and Read, C. (**1992**). *The Acoustic Analysis of Speech* (Singular, San Diego).

Kewley-Port, D., Burkle, T. Z., and Lee, J. H. (**2007**). "Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners," J. Acoust. Soc. Am. **122**, 2365–2375.

Ladefoged, P. (**2001**). *Vowels and Consonants: An Introduction to the Sounds of Languages* (Blackwell, Oxford).

Lee, J. H., and Kewley-Port, D. (**2009**). "Intelligibility of interrupted sentences at subsegmental levels in young normal-hearing and elderly hearing-impaired listeners," J. Acoust. Soc. Am. **125**, 1153–1163.

Leung, H. C., and Zue, V. W. (**1984**). "A procedure for automatic alignment of phonetic transcriptions with continuous speech," in Proceedings of ICASSP'84, pp. 2.7.1–2.7.4.

Li, N., and Loizou, P. (**2007**). "Factors influencing glimpsing of speech in noise," J. Acoust. Soc. Am. **122**, 1165–1172.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). "Perception of the speech code," Psychol. Rev. **74**, 431–461.

Miller, J. L. (**1994**). "On the internal structure of phonetic categories: A progress report," Cognition **50**, 271–285.

Miller, G., and Licklider, J. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.

Owens, E., Talbot, C. B., and Schubert, E. D. (**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.

Ramus, F., Nespor, M., and Mehler, J. (**1999**). "Correlates of linguistic rhythm in the speech signal," Cognition **75**, AD3–AD30.

Read, C. and Schreiber, P. A. (**1982**). "Why short subjects are harder to find than long ones," in *Language Acquisition: The State of the Art*, edited by E. Wanner and L. Gleitman (Cambridge University Press, Cambridge).

Rhebergen, K. S., and Versfeld, N. J. (**2005**). "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. **117**, 2181–2192.

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.

Stevens, K. N. (**2002**). "Toward a model for lexical access based on acoustic landmarks and distinctive features," J. Acoust. Soc. Am. **111**, 1872–1891.

Strange, W. (**1987**). "Information for vowels in formant transitions," J. Acoust. Soc. Am. **26**, 550–557.

Strange, W., Jenkins, J. J., and Johnson, T. L. (**1983**). "Dynamic specification of coarticulated vowels," J. Acoust. Soc. Am. **74**, 695–705.

Tajima, K., and Port, R. (**2003**). "Speech rhythm in English and Japanese," *Papers in Laboratory Phonology VI*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press, Cambridge), pp. 317–334.

Toro, J. M., Nespor, M., Mehler, J., and Bonatti, L. L. (**2008**). "Finding words and rules in a speech stream: Functional differences between vowels and consonants," Psychol. Sci. **19**, 137–144.

Zue, V. W., and Seneff, S. (**1988**). "Transcription and alignment of the TIMIT database," in Proceedings of the Second Meeting on Advanced Man-Machine Interface Through Spoken Language, pp. 11.1–11.10.

Developmental factors and the non-native speaker effect in auditory-visual speech perception[a)]

Yuchun Chen[b)]
Department of Human Communication Science, UCL, Gower Street, London WC1E 6BT, United Kingdom

Valerie Hazan[c)]
Speech, Hearing and Phonetic Sciences, UCL, Gower Street, London WC1E 6BT, United Kingdom

(Received 27 March 2008; revised 9 May 2009; accepted 19 May 2009)

The study investigated language and developmental factors in the use of visual information in audiovisual speech perception in speakers from different language backgrounds. Mandarin-Chinese and English adults and 8- to 9-year-old children were presented with /ba/, /da/, and /ga/ tokens spoken by two English and two Mandarin-Chinese speakers. A syllable identification task was presented in auditory-only, visual-only, and audiovisual (congruent and incongruent) conditions in clear and in noise. The results showed an increase in the use of visual information in adults relative to children in both the Chinese and English groups. In addition, a positive correlation between the total visual effect and speechreading performance was found, suggesting that the smaller visual influence in the bimodal condition for children might be accounted by their less sophisticated speechreading ability. In regard to the language factor, it was found that Chinese perceivers use visual information in their audiovisual speech processing to the same extent as English perceivers. Finally, there was evidence for a "non-native speaker effect" (i.e., stronger visual effect for non-native speech stimuli), but only for the English participants. Results from the current study suggest that the visual appearance of individual speakers and the acoustic-phonetic properties of specific languages should be considered in future cross-language studies.
 [DOI: 10.1121/1.3158823]

PACS number(s): 43.71.Gv, 43.71.Hw, 43.71.Ft [AJ] Pages: 858–865

I. INTRODUCTION

It is well accepted that people use visual (speechreading) information in face-to-face communication. Visual speech facilitates comprehension not only in the presence of background noise (Sumby and Pollack, 1954) but also when auditory signals are clear and intact (Reisberg *et al.*, 1987). Moreover, visual speech appears to influence perception when it is discrepant with auditory speech as shown in the "McGurk effect:" When a visual /ga/ syllable is combined with an auditory /ba/ syllable, subjects report a response (typically /da/) that provides the best fit to the conflicting information (McGurk and MacDonald, 1976).

Since the original report of the McGurk effect, the visual biasing effect on speech perception has been established as a robust effect in English-speaking cultures, but there is conflicting information as to whether this process of early audiovisual integration might be universal or language/culture-specific. Cross-linguistic studies of audiovisual perception are essential to address this question, but there are difficult methodological issues to address when designing such studies. If language-specific audiovisual stimuli are used with different language groups, the comparison of visual weightings across groups is likely to be affected by idiosyncrasies of the stimuli used in each language. If common audiovisual stimuli are used across language groups, the visual effect might be biased by the fact that subjects seem to show greater visual weighting when perceiving a non-native speaker ("foreign language effect," Sekiyama and Tohkura, 1993). Both these approaches have been used in different studies. Using audiovisual stimuli in which a synthesized /ba/-/da/ continuum was synchronized with a computer-animated visual face, Massaro *et al.* (1993) showed that the influence of visible speech appeared to be of the same magnitude for native speakers of Japanese, Spanish, and English. In a later study (Massaro *et al.*, 1995), the same stimuli were used to extend this cross-linguistic research to another language, Dutch. Results in an open-set task showed that although speakers of different languages might process audiovisual speech in the same way, they do give judgments corresponding to phoneme categories that occur in their native language. de Gelder and Vroomen (1992), who measured the contribution of the visual source to bimodal speech events by Mandarin-Chinese and Dutch speakers using Massaro's synthetic stimuli, also found that the mean visual effect did not differ significantly across the Chinese and Dutch groups. In a recent study, Chen and Massaro (2004) used the same synthesized five-level /ba/-/da/ continuum as in previous studies (Massaro *et al.*, 1993, 1995) to examine bimodal

[a)]Portions of this work were presented in "Language Effects on the Degree of Visual Influence in Audiovisual Speech Perception," Proceedings of the 16th ICPhS, Saarbrücken, August 2007, and "Developmental Factor in Auditory-Visual Speech Perception—The McGurk Effect in Mandarin-Chinese and English Speakers," Proceedings of AVSP, Hilvarenbeek, September 2007.

[b)]Electronic mail: yuchun.m.chen@gmail.com
[c)]Present address: Department of Special Education, National Taiwan Normal University, Taiwan.

speech perception of native Mandarin-Chinese speakers, and found a significantly greater proportion of /da/ responses for the Mandarin-Chinese speakers ($M=0.58$) than for the English speakers ($M=0.44$). Although these differences might be due to specific properties of the information input such as differences in phonemic inventories or in the phonetic realizations of the syllables, the authors suggested that the underlying nature of audiovisual speech processing is similar across two languages and could be explained by the same mathematical model of multimodal integration (i.e., fuzzy logical model of perception; Massaro, 1998).

Studies taking the approach of using language-specific stimuli have found somewhat different and sometimes conflicting results. Using natural audiovisual stimuli, Sekiyama and Tohkura (1993) tested Japanese and American participants on language-specific stimuli produced by a single speaker in each language and found that the Japanese perceivers showed a very weak McGurk effect when listening to clear Japanese speech, but a highly-increased effect when noise was added, while American participants showed evidence of a McGurk effect in clear stimuli also. They proposed possible cultural and linguistic factors that might account for the weak McGurk effect of Japanese speakers in their native language. A potential cultural factor was the norm of not directly looking speakers in the face during conversation, especially when the person is of a higher status. They suggested that this cultural habit may cause Japanese speakers to develop a type of processing that gives greater weight to auditory information even in face-to-face communication. To test this hypothesis, Hayashi and Sekiyama (1998) tested Chinese perceivers on the same Japanese and English stimuli, as face-avoidance also occurs in the Chinese culture, and found that the visual effect shown by Chinese participants was similar to the Japanese perceivers for Japanese stimuli and even weaker for the English stimuli. They therefore claimed strong support for the face-avoidance hypothesis.

However, another explanation for the weaker visual effect seen in Japanese perceivers may be linked to characteristics of their phoneme inventory (Sekiyama and Burnham, 2008). Indeed, Japanese contains only five vowels (/a/, /i/, /u/, /e/, and /o/) and fewer consonants and syllable structures than English, with no consonant clusters, resulting in a lower degree of acoustic-phonetic complexity. Furthermore, Japanese has less visually-identifiable elements than English and lacks certain visually-salient consonants such as labiodentals and interdentals (e.g., /f/, /v/, /θ/, and /ð/) that are easy to speechread. Since the acoustic-phonetic characteristics of Japanese are relatively salient and the speechreading information less so, it is suggested that Japanese speakers might develop a more auditory-dependent type of speech processing, with a significant use of visual information only when the auditory speech is indistinct. A weaker McGurk effect in Chinese perceivers might also be due to an auditory bias linked to the tonal characteristics of the Chinese language, given that tones, which can distinguish lexical items, are more easily distinguished acoustically than visually. Japanese is recognized as a pitch-accent language in which two tones are applied to some particular strings of consonants and vowels to create different meanings. Accordingly, if a person's auditory reliance in audiovisual speech perception depends on the extent to which the perceiver's native language is tonal, the reduced McGurk effect in Japanese and Chinese speakers may also be explained solely by the tonal characteristics of their language.

One shortcoming of the studies carried out so far with Chinese speakers is that they have all been carried out using either non-native stimuli [e.g., Japanese and English stimuli in the studies of Sekiyama *et al.* (2003)] or synthetic speech stimuli that are less likely to induce a "native" mode of listening. For a proper test of whether there is evidence of a lesser visual effect in Chinese speakers, it is therefore necessary to test them with stimuli produced by native Chinese speakers, as done in this study.

Another issue of interest regarding visual effects in speech perception is the extent to which these are subject to a developmental trajectory. In their original study, McGurk and MacDonald (1976) included 3- to 5-year-old and 7- to 8-year-old children as well as adults and reported that the two groups of English-speaking children displayed a smaller overall visual influence than adults. Although the three groups all identified auditory-only stimuli correctly, the visually-influenced response rates were 59%, 52%, and 92% for preschoolers, schoolchildren, and adults, respectively. The reduced visual influence in children's audiovisual speech perception is robust, as it has been confirmed in later studies. For instance, Massaro (1984) reported similar weaker visual effects for children than adults using synthesized speech sounds ranging from /ba/ to /da/ combined with a face articulating either /ba/ or /da/. In order to further test a possible explanation of this difference, children and adults were tested with both auditory and visual sources and were required to identify speech events on the basis of the visual source only (Massaro *et al.*, 1986). In the bimodal condition, the differences in the percentages of correct auditory /da/ identifications for a visual /da/ and a visual /ba/ stimulus were 82% for adults and 35% for children. Also, children were found to be poorer speechreaders than adults (visual-only scores of 79% for children vs 96% for adults). A positive correlation was observed between speechreading ability and the size of the visual contribution to bimodal speech perception for both children and adults. The results clearly argue in favor of poorer speechreading abilities leading to lower effect of visual input in bimodal conditions in children relative to adults. Hockley and Polka (1994) also reported a developmental increase across ages (5-, 7-, 9-, 11-year-olds, and adults) in speechreading ability and in the degree of visual influence in audiovisual speech perception using the McGurk paradigm. An adult-like response pattern was observed in only half of the children in the oldest child group, suggesting that the ability to use visual speech information continues to develop beyond the age of 12.

In order to obtain detailed information of the role of linguistic experience in the development of auditory-visual speech processing, Sekiyama *et al.* (2003) combined the developmental and the cross-language McGurk paradigm to investigate the developmental and language factors in audiovisual speech perception, using both Japanese and English

stimuli. Visual influence was low for both Japanese and English 6-year-olds, and increased significantly by adulthood for English but not Japanese participants. This suggests that the inter-language difference in visual effect develops some time between 6 years and adulthood (20 years). In a recent study (Sekiyama and Burnham, 2008), the same paradigm was used to test English and Japanese children aged 6-, 8-, and 12-year-old. While the 6-year-olds showed no difference in response frequency and reaction time, there was a developmental increase in visual influence between 6 and 8 years for English children only, thus resulting in inter-language differences from 8 years onward. Sekiyama and Burnham (2008) attributed these differences to the nature of the surrounding language, such as phonological complexity and visual distinctness of a specific language.

Since previous data on bimodal speech perception for Chinese adults are contradictory and based only on nonnative stimuli, and as there has been no developmental study in Chinese children, the aim of this study was twofold. First, using stimuli produced by both Chinese and English speakers, the authors tested the hypothesis that Chinese adults show a lower visual effect than English speakers, as previously found for Japanese adults. The second aim of the study was to test whether the lack of developmental effect in Japanese children could be replicated for children of another language background (Mandarin-Chinese) that had also been described as being less subject to a visual effect than English.

II. EXPERIMENT

In order to investigate language and developmental factors in audiovisual speech processing in native Chinese speakers, the experimental procedures used by Sekiyama *et al.* (2003) were partly replicated. An English control group was chosen since the audiovisual speech perception of native English speakers is well established from previous studies.

A. Participants

Two adult groups and two child groups participated in this study. Gender was approximately balanced in each group, all participants had normal (or corrected-to-normal) vision, and none reported any history of speech or hearing disorders. The adult groups included 22 Mandarin-Chinese and 18 British English adults (age range: 20–54 years, mean age: 31 years). The Chinese adults were tested in Taiwan where Mandarin is the formal language; none had lived in a foreign country for more than 6 months. For most of the Chinese adult participants, education in English as a foreign language started at age 12 in schools that emphasized grammar and reading rather than oral skills, and the average duration of formal English education was 6 years. The child groups included 20 Mandarin-Chinese children tested in Taiwan and 18 English children tested in the United Kingdom aged between 8 and 10 years (mean age: 8.9 years). This age range was chosen because it was an age range at which a cross-language difference was already significant in Sekiyama and Burnham's (2008) study.

B. Stimuli

Stimuli consisted of the syllables /ba/, /da/, and /ga/ uttered by four speakers recorded in London (two Chinese and two English, one male and one female in each language). The three consonants /b/, /d/, and /g/ and vowel /a/ have phonemic status in both Mandarin-Chinese and English. Chinese speakers were asked to pronounce the syllables with a falling tone (tone 4) in Mandarin-Chinese. For all speakers, video recordings were made in a sound-attenuated room. The speaker's face was set against a blue background and illuminated with key and fill lights. The speaker's head was fully visible within the frame. Video recordings were made to a Canon XL-1 DV camcorder. Audio was recorded from a Bruel and Kjaer type 4165 microphone to both the camcorder and to a DAT recorder. The video channel was digitally transferred to a PC and time-aligned with the DAT audio recording, which was of higher quality than the audio track of the video. Video clips were edited so that the start and end frames of each token showed a neutral facial expression. Stimuli were down-sampled once the editing had been completed (250×300 pixels, 25 f/s, and audio sampling rate 22.05 kHz) after the intensity of all stimuli had been normalized to a fixed level.

Three types of stimuli were prepared: visual-only (V), auditory-only (A), and audiovisual (AV) stimuli. Half of the AV stimuli were audiovisually congruent, and half were audiovisually incongruent. To construct the incongruent stimuli, tokens of /ba/, /da/, and /ga/ were selected that were most similar in terms of their duration, intonation contours, and facial movements. The A and V channels were aligned to ensure that there was auditory-visual coincidence at consonant release. As in Sekiyama and Burnham, 2008, the three incongruent AV stimuli in this study were (1) auditory-ba/visual-ga, (2) auditory-da/visual-ba, and (3) auditory-ga/visual-ba. The V stimuli were created by cutting out the audio track; in the A stimuli, the visual track was replaced with a still face of the speaker with mouth closed. In order to test whether visual influence is greater when the auditory signal is less clear, a "noisy" condition was included for AV and A conditions. Pink noise was added to the auditory channel using SFS software (Speech Filing System, Huckvale, 2004) at a signal-to-noise ratio (SNR) of -12 dB. This SNR level was selected following a short pilot test in which different levels were presented to one English and one Chinese listener. Only two conditions (clear and -12 dB) were included since the noise factor was not the focus of the current study.

Experimental conditions were blocked according to the modality (A, V, and AV) and the SNR of the auditory stimuli (clear, -12 dB), with four repetitions of each stimulus per block. Thus, participants received 48 trials in each of the A, A-noisy, and V blocks (3 consonants $\times 4$ speakers $\times 4$ repetitions), and 96 in each of the AV-clear and AV-noisy conditions (3 auditory consonants $\times 2$ congruity types $\times 4$ speakers $\times 4$ repetitions). There were five AV practice trials before the whole test began and the total number of trials per participant was 336.

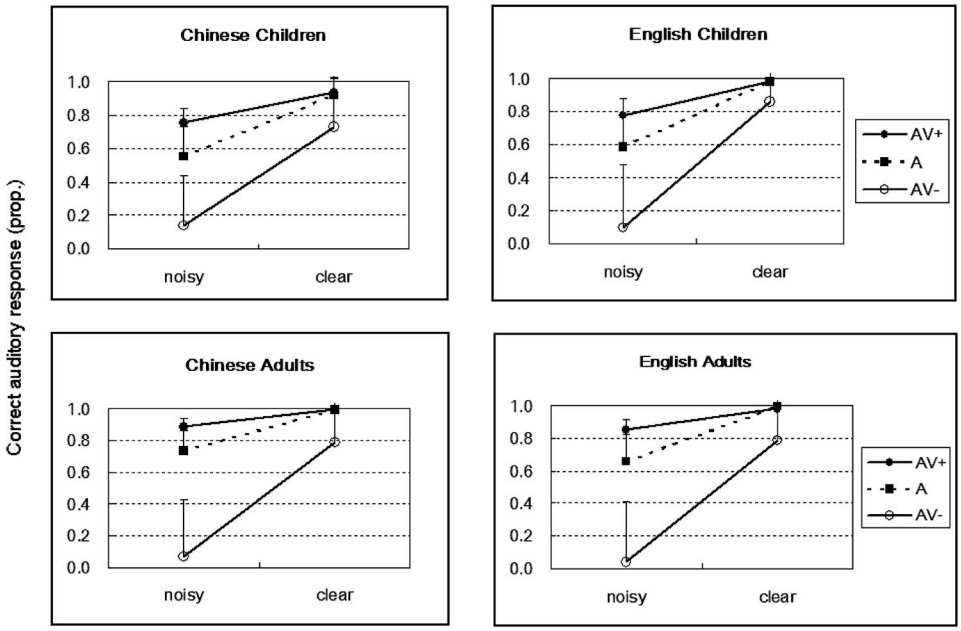


FIG. 1. Proportion of correct auditory responses in AV-congruent (AV+), AV-incongruent (AV–), and auditory (A) conditions in each group. Error bars show 1 standard error (s.e.).

C. Procedure

The testing software was designed using the CSLU toolkit (e.g., Cole *et al.*, 1999). The video stimuli with the speaker's head fully visible were presented in a 3×3 in.^{2} frame on the color monitor of a laptop, and the auditory channel was presented to both ears at a comfortable listening level through headphones. Since it was found in Jordan and Sergeant, 1998 that head height sizes of 42 mm did not lead to a reduction in visual effects relative to full size heights, the small frame used in this study was not expected to affect the degree of visual bias, and had the advantage of providing enough space for arranging the three response labels on the screen. All participants were tested with the same apparatus individually in a quiet room in Taiwan or in England. Conditions were presented in the following order: AV, AV-noisy (AVn), A, A-noisy (An), and V blocks. The V condition was given last because of its difficulty and the A condition was interposed to avoid a transfer of visual effect from AV to V block. Within each block, stimuli were presented in random order. The test was response-paced, with breaks after every 30 trials in the AV and AVn blocks.

In the AV condition, participants were instructed to click on one of three labels on the screen (BA, DA, and GA) to answer what they had heard while looking at and listening to each syllable. For the Chinese children, three additional "Zhuyin Fuhao" symbols (the national Chinese phonemic system used in Taiwan) equivalent to the three syllables were stuck next to the three corresponding buttons on the screen to help their understanding. The classical /ða/ choice was not included since /ð/ is not a valid consonant label in Mandarin-Chinese. As in Sekiyama *et al.*, 2003, "combination responses," e.g., "bda" or "bga," were also not allowed in order to make the response alternatives less confusing for young children. In the A condition, the participants' task was to answer only what they had heard. In the V condition, they were asked to read lips and click on the label they thought the speaker was pronouncing.[1] During the whole test, which lasted around 25 min for adults and 35 min for children, the experimenter sat next to the participants to make sure they were watching the screen at all times.

D. Results and discussion

1. Degree of visual effect

As in Sekiyama *et al.*, 2003, the positive effect of visual information was described as the difference between auditory accuracy in the congruent AV stimuli (AV+) and in the A stimuli, and negative (interference) effect was calculated as the difference between auditory accuracy in the incongruent AV stimuli (AV–) and A stimuli. For the main analyses, these measures were calculated per participant across all four speakers, as both perceiver groups heard the same number of stimuli spoken by native and non-native speakers (see Fig. 1). The total visual effect was measured by combining the positive and negative visual effects (i.e., the difference of correct auditory response between AV+ and AV– conditions), and this measure was used in the following analyses.

Group difference in the size of the visual effect was analyzed via a repeated-measure ANOVA, with age and participant language as across-subject factors and noise condition as within-subject factor. The main effects of age ($F[1,74]=20.275$, $p<0.0001$) and noise ($F[1,74]=683.041$, $p<0.0001$) were highly significant. The interaction effect between age and noise factors ($F[1,74]=9.871$, $p=0.002$) suggests greater visual influence for adults in the noisy condition. However, there was no significant main effect of language background ($F[1,74]=0.223$, $p=0.639$) so no evidence of a lower degree of visual influence for Chinese participants than English participants. To identify the source of age factor, two separate ANOVAs were further conducted for clear and noisy conditions. The source of difference for age was from the noisy condition ($F[1,74]=38.406$, $p<0.001$), with no group differences in the clear condition ($F[1,74]=1.279$, $p=0.262$). Thus, these results suggest that

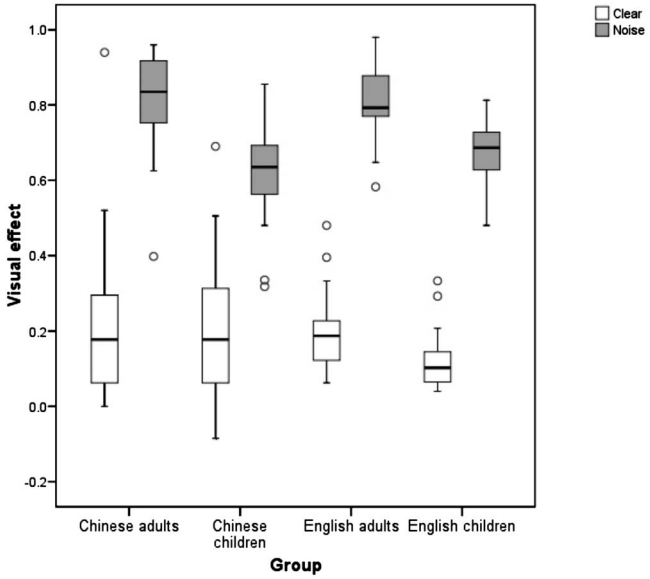


FIG. 2. Mean proportion of visual effect (different between auditory accuracy in AV+ and AV–) across four groups under clear and noisy conditions. Circles represent outliers with values between 1.5 and 3 s.e.

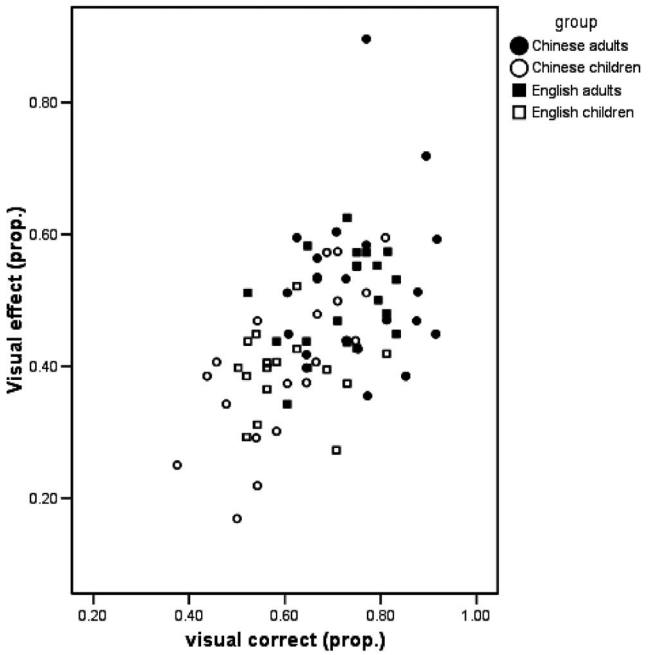


FIG. 3. Correlation between the proportion of visual correct response and the size of total visual effect for all participants.

the degree of visual influence was the same for Chinese and English children at age 8 as well as in adulthood. There was an increase in visual effect in adults relative to children in both the Chinese and English groups but only in the noisy condition (see Fig. 2).

2. A-only and V-only conditions

The percentages of correct responses were calculated in the A and V conditions (see Table I). In the A-only condition, repeated-measure ANOVAs showed that higher percentages of correct responses were obtained in the clear than noisy conditions ($F[1,74]=1831.574$, $p<0.001$). The age-related difference was significant ($F[1,74]=58.193$, $p<0.001$): Adults showed better performance than children in identifying purely auditory signals, especially in the noisy condition (age × noise interaction: $F[1,74]=27.405$, $p<0.001$).

Similarly, in the V-only (speechreading) condition, the results revealed an age-related difference only ($F[1,74]=31.325$, $p<0.001$), suggesting a developmental increase in speechreading ability. A positive correlation was observed between speechreading ability (score in V condition) and the degree of overall visual effect ($r=0.536$, $p=0.01$; see Fig. 3). The pattern of results confirms that the less sophisticated speechreading ability in children might be the reason for the poorer performance in the bimodal conditions as suggested in Massaro, 1984.

3. "Non-native speaker" effect

It was proposed in previous cross-language studies that the visual effect was stronger for non-native speech stimuli than for native speech stimuli (e.g., Sekiyama and Tohkura, 1993; Sekiyama *et al.*, 1995; Fuster-Duran, 1996). Since both groups of participants perceived stimuli spoken by both native and non-native speakers, the next analysis examined whether there was a non-native speaker effect across language and age groups. In the calculation of the total visual effect, the stimuli were further divided according to the two speaker groups, i.e., Chinese and English. Combining the visual effect under clear and noisy conditions, all participants showed a greater visual effect when the stimuli were Chi-

TABLE I. Percent correct scores for Chinese (Chi) and English (Eng) speaker groups in the auditory-only (A) (in clear and in noise) and visual-only (V) conditions for the four perceiver groups. Standard deviant measures are given in parentheses.

	Condition								
	Speaker								
	A clear			A noise			V		
Perceiver	Chi	Eng	Total	Chi	Eng	Total	Chi	Eng	Total
Chinese children ($n=20$)	94.4(6.5)	88.6(10.5)	91.5(8.1)	63.6(11.0)	45.8(8.5)	54.7(8.4)	61.4(15.2)	60.0(12.0)	60.7(12.0)
Chinese adults ($n=22$)	99.5(1.4)	99.1(2.1)	99.3(1.6)	84.9(10.9)	61.4(9.7)	73.2(6.7)	74.1(10.1)	76.5(13.4)	75.3(10.2)
English children ($n=18$)	97.9(4.3)	98.9(2.3)	98.3(2.2)	55.7(10.8)	60.2(9.4)	57.9(6.7)	57.3(11.3)	64.9(12.7)	61.1(9.2)
English adults ($n=18$)	99.1(2.3)	100(0.0)	99.5(1.1)	64.8(8.3)	65.7(11.1)	65.3(8.1)	65.3(9.9)	79.9(11.8)	72.6(9.1)
All ($N=78$)	97.7(4.5)	96.6(7.2)	97.1(5.4)	68.1(15.1)	58.1(12.1)	63.1(10.4)	64.9(13.2)	70.4(14.8)	67.7(12.1)

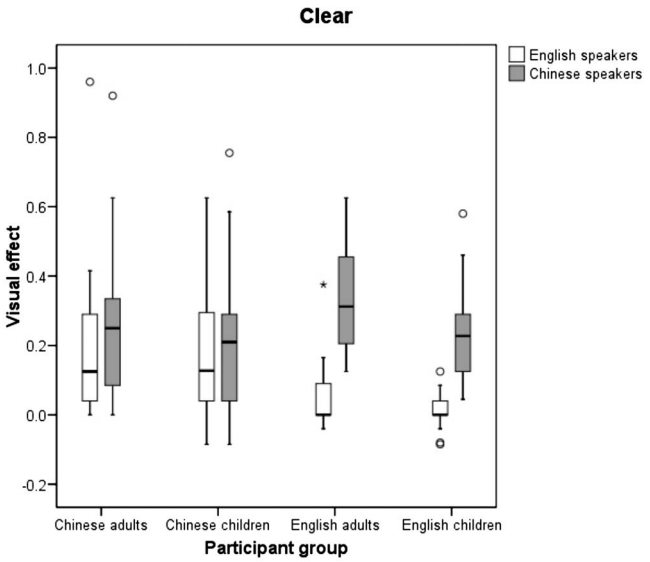


FIG. 4. Comparison of the proportion of visual effect between the Chinese and English speaker groups for the four perceiver groups in clear condition. Asterisks represent extremes with values greater than 3 s.e.

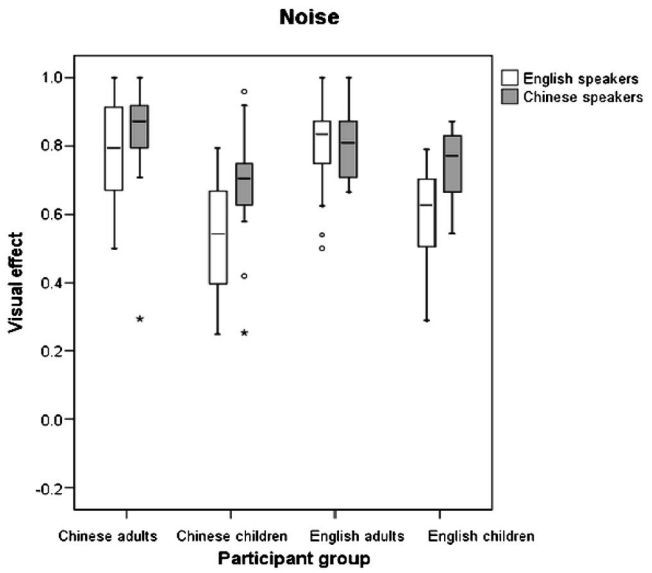


FIG. 5. Comparison of the proportion of visual effect between the Chinese and English speaker groups for the four perceiver groups in noisy condition.

nese, which might be due to individual characteristics of the two Chinese speakers. The difference in visual effect for the Chinese and English speakers was significantly greater for English participants ($F[1,74]=17.793$, $p<0.001$), suggesting that English children and adults were more influenced by visual information when the speakers were non-native (see Figs. 4 and 5). There was no evidence of a greater magnitude of this non-native speaker effect in children than adults.

In summary, an evident increase in the use of visual information between the 8- and 9-year-old and adult groups was found in both Chinese and English groups in noisy conditions. In the A-only and V-only conditions, adults also showed overall a better performance than children, irrespective of their language background. In addition, as in previous findings by Massaro (1984), the authors' results provide further evidence that the smaller influence of visual information in the bimodal condition for children might be accounted for by their less sophisticated speechreading ability.

III. GENERAL DISCUSSION

The current study aimed to provide a better understanding of the role of language experience in the development of audiovisual speech processing. First, it examined whether the lower visual effect in Japanese adults and the lack of developmental effect in Japanese children seen in Sekiyama *et al.*, 2003 could be replicated for speakers of another language background (Mandarin-Chinese) that has also been described as being less subject to a visual bias than English. It was expected that Mandarin-Chinese perceivers would show lower effect of visual information according to a number of hypotheses, including cultural and linguistic factors (Sekiyama, 1997; Sekiyama *et al.*, 2003). Nevertheless, a direct comparison of results for Chinese and English adults in the current study showed inconsistent evidence. When listening to native speakers of their language, contrary to expectations, Chinese perceivers showed a greater degree of visual influence than did English perceivers and both groups scored similarly in the A- and V-only conditions. The inconsistent result challenges the possible cultural and linguistic hypotheses. The cultural habit of face-avoidance is less observed in the younger Japanese generation due to the influence of western cultures (Isei-Jaakkola, 2006) and there seems to be a similar trend in the Taiwanese society. Since this face-avoidance habit might be more prevalent in the older generation, this reduction in face-avoidance culture might partly explain the variation in Chinese adults' performance between previous studies and the current study. As for the linguistic factor, the tonal hypothesis (Sekiyama *et al.*, 2003) suggests that speakers of tone languages would show a lower degree of visual influence, as auditory cues are more informative for tone perception than visual cues. Although it may be true that auditory information is more informative in identifying tones, it does not necessarily imply that there is no visual information for lexical tone. In fact, recent studies have shown that there are some articulatory changes for different tones in Mandarin-Chinese, especially for tone 3 (Erickson *et al.*, 2004; Hoole and Hu, 2004). Recent studies have also shown that perceivers may be able to distinguish tones to some extent based on visual information alone even when their native language is not tonal (Australian English speakers) (Burnham *et al.*, 2001). Furthermore, Chen and Massaro (2008) found that Mandarin-Chinese participants from Mainland China or Taiwan who were taught specific identification strategies could use the speakers' neck activities and head/chin movements to discriminate tones in Mandarin-Chinese, especially tone 1 and tone 3. Since both tonal and non-tonal native speakers seem to use some visual information in identifying tone variations and the tonal information did not induce greater auditory reliance in the McGurk effect, the tone hypothesis still awaits further investigation.

Another possible linguistic explanation for the result in the current study of a similar visual effect between English

and Chinese participants is the fact that Chinese has a rich phonetic inventory. Unlike Japanese, which has only 5 vowels, there are 16 vowels in Mandarin-Chinese. In terms of consonant inventories, there are 21 consonants in Mandarin-Chinese. Although there are no consonant clusters, it has visually-identifiable labio-dental /f/, which is easy to lip-read. In addition, there are two sets of syllable onsets that are identified by the retroflex and lip-rounding properties (i.e., /jhih, chih, shih, rih/ vs /zih, cih, sih/).[2] Since speechreading information is useful for Mandarin-Chinese speakers under certain circumstances, the assumption that Mandarin-Chinese speakers possess a more auditory-dependent type of speech processing than English speakers (Sekiyama, 1997) might need to be checked further. In fact, in a study examining how native language affects audiovisual perception of non-native speech, Chinese participants relied more on visual information for unfamiliar sounds compared to Korean subjects (Wang *et al.*, 2007). Sekiyama and Burnham (2008) also suggested that a broader language environment, rather than the single tonal hypothesis, might be the crucial factor in the developmental onset of inter-language differences between Japanese and English auditory-visual speech perception.

The "non-native speaker effect" was another issue examined in this study. The hypothesis is that people use more visual information (as shown by a greater visual effect) when the stimuli are non-native (e.g., Sekiyama and Tohkura, 1993), even if the speakers are producing syllables that occur in the phoneme inventory of the listeners' language. In the current study, it was found that when the stimuli were presented audiovisually, English perceivers used significantly more visual information when the stimuli were Chinese, even though they were poorer at speechreading Chinese speakers in the visual-only condition. This therefore shows evidence of a non-native speaker effect. However, Chinese native perceivers did not show such a difference between the two language speaker groups, which countered this hypothesis. The lack of non-native speaker effect in Chinese participants is consistent with Hayashi and Sekiyama's (1998) study. A possible explanation is that while none of the English participants had learned Chinese, all the Chinese participants had knowledge of English through the classes they had taken for at least 6 years in school. In addition, they have many opportunities to watch English speakers on television or other media. Furthermore, in Taiwan, many English teachers in bilingual schools or after-school classes are native English speakers. Thus, this familiarity for English speakers might account for the lack of non-native speaker effect in Chinese participants. Chinese participants showed much better auditory accuracy for Chinese tokens in the A-only condition. Chinese participants may have shown differences between the native and non-native stimuli in the A-only condition because of the tonal characteristic of Mandarin-Chinese. Since the Chinese stimuli in the current study were pronounced using tone 4 in Mandarin and were real words in Chinese, it might be easier for Chinese speakers to distinguish the three syllables. It should be noted that one limitation of the current study regarding the non-native speaker effect hypothesis is the fact that only two speakers were included in each language group. Thus, to investigate the non-native speaker effect in detail, future studies should include more speakers in each language group in order to reduce the impact of individual speaker variations. Indeed, given that McGurk studies have typically used small numbers of speakers, discrepancies across studies may be due to differences in the visual clarity of individual speakers. Visual weighting may also be affected by methodological issues such as the choice of headphone over loudspeaker presentation. If visual weighting can so easily be changed as a result of individual speaker characteristics or differences in methodology, however, it might be possible that these considerations would be a more fundamental reason for the variation in audiovisual speech perception seen across audiovisual studies than language-related factors.

Finally, an evident developmental increase in the use of visual information was found in both Chinese and English groups in noisy conditions, which is in accordance with previous and recent studies for English participants (e.g., Sekiyama and Burnham, 2008). In addition, in the A-only and V-only conditions, adults also showed overall better performances than child groups, irrespective of their language background. As in previous findings by Massaro *et al.* (1986), a positive correlation between the total visual effect and speechreading performance was found. Therefore, the authors' results provide further evidence that the smaller influence of visual information on the bimodal condition for children might be accounted by their less sophisticated speechreading ability. In sum, the current study suggests that there is a general tendency for visual speech processing to increase over age for both Chinese and English subjects, and this speechreading ability might facilitate the acquisition of audiovisual speech perception.

In conclusion, the current study suggests that Mandarin-Chinese native speakers are no different from English native speakers in their use of visual information in auditory-visual speech perception at the age of 8 as well as in adulthood. The degree of visual bias in the two groups does not seem related to face-avoidance or tonal aspects of language. In regard to the non-native speaker effect, only the English perceivers showed stronger visual effect for the non-native stimuli. Whether this non-native speaker effect is related to individual speaker characteristics, to the assumed non-nativeness of the speakers based on their visual appearance or their acoustic-phonetic properties of the stimuli still awaits further investigation.

[1]Although the different instructions given for each condition might have introduced a bias, the same stimuli were used in a later study (Hazan and Li, 2008) in that subjects were asked to "decide which syllable they perceived" after each presentation and the same general visual effect for the AV stimuli was obtained.

[2]The sounds are represented by Tongyong Pinyin, one of the Romanization systems of the Chinese language used in Taiwan.

Burnham, D., Lau, S., Tam, H., and Schoknecht, C. (**2001**). "Visual discrimination of Cantonese tone by tonal but non-Cantonese speakers, and by non-tonal language speakers," in Proceedings of the International Conference on Auditory-Visual Speech Processing, Sydney, Australia, pp. 155–160.

Chen, T. H., and Massaro, D. W. (**2004**). "Mandarin speech perception by

ear and eye follows a universal principle," Percept. Psychophys. **66**, 820–836.
Chen, T. H., and Massaro, D. W. (**2008**). "Seeing pitch: Visual information for lexical tones of Mandarin-Chinese," J. Acoust. Soc. Am. **123**, 2356–2366.
Cole, R., Massaro, D. W., de Villiers, J., Rundle, B., Shobaki, K., Wouters, J., Cohen, M., Beskow, J., Stone, P., Connors, P., Tarachow, A., and Solcher, D. (**1999**). "New tools for interactive speech and language training: Using animated conversational agents in the classroom of profoundly deaf children," Proceedings of the ITRW on Methods and Tools in Speech Science Education (MATISSE), London, pp. 45–52
de Gelder, B., and Vroomen, J. (**1992**). "Auditory and visual speech perception in alphabetic and non-alphabetic Chinese/Dutch bilinguals," in *Cognitive Processing in Bilinguals*, edited by R. J.Harris (Elsevier, Amsterdam), pp. 413–426
Erickson, D., Iwata, R., Endo, M., and Fujino, A. (**2004**). "Effect of tone height on jaw and tongue articulation in Mandarin Chinese," in Proceedings of TAL-2004, Beijing, China, pp. 53–56.
Fuster-Duran, A. (**1996**). "Perception of conflicting audio-visual speech: An examination across Spanish and German," in *Speechreading by Humans and Machines*, edited by D. G. Stork and M. E. Hennecke (Springer-Verlag, Berlin), pp. 135–143
Hayashi, Y., and Sekiyama, K. (**1998**). "Native-foreign language effect in the McGurk effect: A test with Chinese and Japanese," in Proceedings of the International Conference on Auditory-Visual Speech Processing, Sydney, Australia, pp. 61–66.
Hazan, V., and Li, E. (**2008**). "The effect of auditory and visual degradation on audiovisual perception of native and non-native speakers," in Proceedings of the Interspeech 2008, Brisbane, Australia, pp. 1191–1194.
Hockley, N., and Polka, L. (**1994**)."A developmental study of audiovisual speech perception using the McGurk paradigm," J. Acoust. Soc. Am. **96**, 3309.
Hoole, P., and Hu, F. (**2004**). "Tone-vowel interaction in standard Chinese," in Proceedings of TAL-2004, Beijing, China, pp. 89–92.
Huckvale, M. (**2004**). Speech filing system software, version 4.7
Isei-Jaakkola, T. (**2006**)."Cognition and physio-acoustic correlates—Audio and audio-visual effects of a short English emotional statement: On JL2, FL2 and EL1," in Proceedings of the 5th International Conference on NLP, FinTAL, Turku, Finland pp. 161–173.
Jordan, T. R., and Sergeant, P. C. (**1998**). "Effects of facial image size on visual and audiovisual speech recognition," in *Hearing by Eye II: Advances in the Psychology of Speechreading and Auditory-Visual Speech*, edited by R. Campbell, B. Dodd, and D. Burnham (Psychology, London), pp. 155–176
Massaro, D. W. (**1984**). "Children's perception of visual and auditory speech," Child Dev. **55**, 1777–1788.
Massaro, D. W., Cohen, M. M., and Smeele, P. M. (**1995**). "Cross-lingustic comparisons in the integration of visual and auditory speech," Mem. Cognit. **23**, 113–131
Massaro, D. W. (**1998**). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle* (MIT, Cambridge, MA).
Massaro, D. W., Cohen, M. M., Gesi, A., Heredia, R., and Tsuzaki, M. (**1993**). "Bimodal speech perception: An examination across languages," J. Phonetics **21**, 445–478.
Massaro, D. W., Thompson, L. A., Barron, B., and Laren, E. (**1986**). "Developmental changes in visual and auditory contributions to speech perception," J. Exp. Child Psychol. **41**, 93–113.
McGurk, H., and MacDonald, J. W. (**1976**). "Hearing lips and seeing voices," Nature (London) **264**, 746–748.
Reisberg, D., McLean, J., and Goldfield, A. (**1987**). "Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli," in *Hearing by Eye: The Psychology of Lip-Reading*, edited by B. Dodd and R. Campbell (Lawrence Erlbaum Associates Ltd., London), pp. 97–113
Sekiyama, K. (**1997**). "Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects," Percept. Psychophys. **59**, 73–80.
Sekiyama, K., Braida, L. D., Nishino, K., Hayashi, M., and Tuyo, M. (**1995**). "The McGurk effect in Japanese and American perceivers," in Proceedings of the 13th International Congress of Phonetics Sciences, edited by K. Elenius and P. Branderud, pp. 214–217.
Sekiyama, K., and Burnham, D. (**2008**). "Impact of language on development of auditory-visual speech perception," Dev. Sci. **11**, 306–320.
Sekiyama, K., Burnham, D., Tam, H., and Erdener, D. (**2003**). "Auditory-visual speech perception development in Japanese and English speakers," in Proceedings of the International Conference on Auditory-Visual Speech Processing, St. Jorioz, France, pp. 61–66.
Sekiyama, K., and Tohkura, Y. (**1993**). "Inter-language differences in the influence of visual cues in speech perception," J. Phonetics **21**, 427–444.
Sumby, W. H., and Pollack, I. (**1954**). "Visual contribution to speech intelligibility in noise," J. Acoust. Soc. Am. **26**, 212–215.
Wang, Y., Behne, D., Jiang, H., and Feehan, A. (**2007**). "Effect of native language experience on audio-visual perception of English fricatives by Korean and Mandarin natives," in Proceedings of the International Conference on Auditory-Visual Speech Processing, Hilvarenbeek, 1–3 September.

Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers[a)]

Paul Iverson and Bronwen G. Evans
Division of Psychology and Language Sciences, University College London, Chandler House, 2 Wakefield Street, London WC1N 1PF, United Kingdom

(Received 13 August 2008; revised 11 March 2009; accepted 12 May 2009)

This study investigated whether individuals with small and large native-language (L1) vowel inventories learn second-language (L2) vowel systems differently, in order to better understand how L1 categories interfere with new vowel learning. Listener groups whose L1 was Spanish (5 vowels) or German (18 vowels) were given five sessions of high-variability auditory training for English vowels, after having been matched to assess their pre-test English vowel identification accuracy. Listeners were tested before and after training in terms of their identification accuracy for English vowels, the assimilation of these vowels into their L1 vowel categories, and their best exemplars for English (i.e., perceptual vowel space map). The results demonstrated that Germans improved more than Spanish speakers, despite the Germans' more crowded L1 vowel space. A subsequent experiment demonstrated that Spanish listeners were able to improve as much as the German group after an additional ten sessions of training, and that both groups were able to retain this learning. The findings suggest that a larger vowel category inventory may facilitate new learning, and support a hypothesis that auditory training improves identification by making the application of existing categories to L2 phonemes more automatic and efficient.
 [DOI: 10.1121/1.3148196]

PACS number(s): 43.71.Hw, 43.71.Es [RSN] Pages: 866–877

I. INTRODUCTION

One could imagine that the task of learning a second-language (L2) vowel system would be fundamentally different for adults with small and large native-language (L1) vowel systems. For example, novice learners are thought to apply their existing L1 categories to perceive L2 phonemes (e.g., Best, 1995; Best *et al.*, 2001; Trubetzkoy, 1969). This can create ambiguity for individuals who have a small number of L1 vowels, because there are likely to be situations when multiple L2 vowels are assimilated into the same L1 category, making them sound the same (e.g., Spanish speakers hearing English /i/ and /ɪ/ as the same as Spanish /i/; Escudero and Boersma, 2004; Flege *et al.*, 1997; Iverson and Evans, 2007; Morrison, 2002). Moreover, individuals with smaller vowel inventories may use fewer dimensions to distinguish L1 vowels (e.g., only F1 and F2), and need to become sensitive to other aspects (e.g., quantity, diphthongalization, and nasalization) to distinguish vowels in the L2 (e.g., see Bohn, 1995; Bohn and Flege, 1990; Gottfried and Beddor, 1988; Iverson and Evans, 2007; McAllister *et al.*, 2002).

Individuals with large and complex L1 vowel systems may likewise have an early advantage in L2 vowel perception, but their large numbers of categories may make further learning difficult. For example, Flege's (1995, 2003) Speech Learning Model (SLM) claims that L1 and L2 vowels exist in the same phonological space, and learning a new vowel is harder when it is close to an existing category. Individuals with larger L1 vowel systems may therefore have a relatively crowded vowel space that interferes with the formation of new categories, whereas those with smaller L1 vowel spaces may have more room to learn, although it is not clear whether individuals with fewer categories actually have more "uncommitted" vowel space (e.g., Meunier *et al.*, 2003). Moreover, individuals with smaller vowel inventories likely have more incentive to learn given that they have more initial difficulties with L2 vowels.

The available evidence, however, suggests that individuals with large and small L1 vowel systems may learn L2 vowel systems similarly. The types of L1-L2 interactions described above have been well established for individual vowel contrasts such as /i/-/ɪ/ (e.g., Flege *et al.*, 1997; Flege *et al.*, 2003), but we recently found that these kinds of local interactions did not produce fundamentally different ways in which individual Spanish, French, German, and Norwegian listeners learn the English vowel system (Iverson and Evans, 2007). Our study took an individual difference approach, comparing English vowel perception and category representations among listeners with a wide range of L1s, but did not examine learning within individuals (e.g., training or longitudinal study). There were large overall differences in how accurately the language groups recognized English vowels, with lower scores for listeners with smaller L1 vowel systems (i.e., Spanish and French) and higher scores for those with larger L1 vowel systems (i.e., German and Norwegian). However, the acoustic cues that they used were the same; all

[a)]Part I: Iverson, P., Evans, B. G. (**2007**). "Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration," J. Acoust. Soc. Am. **122**, 2842–2854.

groups relied on primary acoustic cues such as F1/F2 target formant frequencies, as well as more fine-grained cues such as formant movement and duration, even though Spanish and French vowels do not contrast in formant movement and duration whereas German and Norwegian vowels do (see also Bohn, 1995).

Moreover, our results also suggested that individuals with small and large L1 vowel systems both had learned aspects of the English vowel inventory (Iverson and Evans, 2007). The subjects completed a vowel space mapping task in which they found best exemplars for vowels in their L1 and L2 (English), and all language groups had systematic differences between their L1 and L2 vowels. For example, the Spanish vowel space had five best exemplars with little formant movement or duration contrast, but the L1 Spanish speakers chose best exemplars for English that were markedly different, with formant movement and duration contrast, as well as a larger number of distinct categories. The Norwegian vowel space was larger (22 vowels), but there were still differences in the vowels that they chose for English; their English /ɪ/ vowel, for example, had a spectrum like that of native English speakers even though Norwegians assimilate English /ɪ/-/i/ into a Norwegian /i/-/i:/ contrast that is made purely with duration. All groups (Spanish, French, German, and Norwegian) exhibited similar amounts of learning, even though L1 assimilation judgments indicated that this learning was not completely necessary for Germans and Norwegians. That is, nearly all English vowels were assimilated into a unique L1 category in German and Norwegian, so these listeners could have, in theory, simply used their existing L1 vowels when listening to English. We thus found no evidence that the larger L1 vowel spaces interfered with new learning.

Such cross-language comparisons are difficult because one cannot completely match the learning experience of different subject groups. For example, even if one could find Spanish and German speakers with identical amounts and ages of English classroom instruction, they could differ, for example, in the type of instruction they received, their exposure to English outside of the classroom, and their individual motivations to learn. The approach of the present study was to control for experience by giving both groups the same amount of auditory training, to further compare the learning of English vowels by individuals with L1 vowel systems that are small and large; Spanish has five vowels with no diphthongs or duration contrast (e.g., Delattre, 1965; Flege, 1989; Stockwell and Bowen, 1965) whereas German has 18 vowels with diphthongs and duration contrast (e.g., Delattre, 1965; Strange *et al.*, 2005). The aim was to examine how their vowel recognition accuracy, L1 assimilation, and vowel space mapping differed before and after training, in order to evaluate whether the Spanish and German L1 vowel spaces made these individuals learn differently.

Several recent studies have adapted the high-variability phonetic training method (Logan *et al.*, 1991) to vowel stimuli, for the purpose of training Japanese adults on English vowels (Lambacher *et al.*, 2005; Nishi and Kewley-Port, 2007), English adults on the Japanese vowel-length contrast (Hirata *et al.*, 2007; Tajima *et al.*, 2008), and English adults on German non-low vowels (Kingston, 2003). Training Japanese adults on monophthongal English vowels has been successful, improving performance by 16–25 percentage points (Lambacher *et al.*, 2005; Nishi and Kewley-Port, 2007; see also Kingston, 2003), whereas training English adults on Japanese vowel-length contrasts has generally yielded smaller degrees of improvement that do not always generalize to untrained phonetic contexts and speaking rates (Hirata *et al.*, 2007; Tajima *et al.*, 2008). Most training protocols have trained listeners using closed-set responses (e.g., long vs short) and small numbers of vowels (e.g., 5). However, Nishi and Kewley-Port (2007) suggested that training on larger sets is more effective overall than concentrating on only the most difficult vowels; Japanese adults who were trained on a set of nine English vowels had broad improvement for all nine vowel categories, but those that were trained on only the three most difficult vowels improved only for these three vowels. The present study trained adults on an even larger set: 14 English vowels including diphthongs.

Previous work on English vowel training has embedded the vowels in CVC contexts (Lambacher *et al.*, 2005; Nishi and Kewley-Port, 2007). The present study instead trained listeners on real English minimal-pair words, to increase the range of phonetic variability and the naturalness of the training materials. Very few minimal pairs can be found that can span the set of 14 English vowels, so we divided this vowel space into four subsets based on cluster analyses of previous vowel confusion data by L2 speakers of English (Iverson and Evans, 2007). For example, listeners could hear the word *pet* and be given the response alternatives *pet, part, pat*, and *putt*. They were thus trained on a relatively large set of vowels, but were given response alternatives that were restricted to a subset of words that they would be expected to confuse.

Training improvements can be difficult to compare across different performance levels because we do not fully understand the underlying mechanisms. For example, there is no way of knowing whether a subject who improves from 20% to 40% recognition accuracy has actually learned the same amount as an individual who improves from 70% to 90%, because we do not know exactly what people are learning and how this translates to identification accuracy. In order to avoid this issue, we selected Spanish and German speakers so that they were matched in terms of their pre-test English vowel identification accuracy. Spanish speakers would normally be expected to be worse than Germans at English vowels, given their small L1 vowel system (Iverson and Evans, 2007). To help equalize this difference, we tested Spanish speakers in London (i.e., regular exposure to English) and tested German speakers in Germany who had little exposure to English outside of the classroom and media. The groups thus differed somewhat in English exposure, but were the same in terms of how well they recognized English vowels.

Both groups of subjects were given the same battery of pre and post tests. They were tested on /b/-V-/t/ words and talkers that were not part of the training set, in order to evaluate their degree of training improvement. Subjects were also tested in terms of L1 assimilation because it is thought that novice learners, at least, perceive vowels in terms of

their native-language phonology (e.g., Best, 1995); we wished to evaluate whether these assimilation patterns could predict recognition accuracy as well as explain improvements in training. Subjects were additionally tested using a vowel space mapping procedure, in which they found best exemplars for English vowels in a large 5-dimensional vowel space that included F1 and F2 target frequencies, F1 and F2 formant movement, and duration (see Iverson and Evans, 2007; Iverson *et al.*, 2006). This evaluated how their underlying notions of what vowels sound good in English changed with training; our previous work demonstrated that individuals whose best exemplars are closer to those of L1 English speakers are also more accurate at recognizing natural recordings of English vowels. Experiment 1 evaluated performance immediately after training. Experiment 2 evaluated retention, as well as the effect of additional training.

II. EXPERIMENT 1: AUDITORY-PHONETIC TRAINING

A. Method

1. Subjects

A total of 33 subjects were initially tested (17 Spanish and 16 German). Pre-test English vowel identification accuracy (see Procedure, Sec. II A 4) ranged somewhat lower for Spanish speakers (30%–83%) than for Germans (42%–89%). In order to match the groups on this measure, the three highest-accuracy German subjects (scores of 87%–89%) and four of the five lowest-accuracy Spanish subjects (scores of 30%–50%) were dropped from the data analysis, creating two groups of 13 subjects each; one relatively low-accuracy Spanish subject (41%) was retained in the study to provide a match for the lowest-accuracy German subject (42%). In these matched groups, the ranges of identification accuracy scores were 41%–83% (mean 67%) for Spanish speakers and 42%–86% (mean 68%) for German speakers. This matching made the interpretation of group differences clearer, but it should be noted that the significant statistical differences reported here (see Results, Sec. II B) remained significant even when all of the original 33 subjects were included.

The Spanish subjects in the matched group were all tested in London and had 1–72 months of experience living in English-speaking countries (median 18 months). They were 21–40 years old (median 27 years), and began learning English when they were 6–34 years old (median 14 years). They came from several countries (Spain, Mexico, Columbia, Peru, Ecuador, Cuba, and Venezuela) but all had a standard Spanish five-vowel system.

The German subjects in the matched group were tested in Potsdam, Germany, and none had lived in English-speaking countries. They were 19–38 years old (median 25 years), and began learning English when they were 9–15 years old (median 12 years). The subjects were predominantly from the Brandenburg region of Germany, and none had non-standard German vowel systems.

The two groups were thus quite different in terms of the length of experience living in English-speaking countries. They were also slightly different in terms of the age of first instruction (median 12 years for German and 14 years for Spanish), even though their median duration of English use (age at test minus age of first instruction) were the same (13 years). In order to evaluate whether these differences could affect the extent that listeners benefited from training, the degree of training improvement (post- minus pre-test identification accuracy; see Results, Sec. II B) was compared using Pearson correlations to the experience and age of first instruction measures, separately for each language group. None of these correlations was significant, $p > 0.05$, suggesting that the age at which the subjects began learning English or the amount of time they spent living in English-speaking countries did not substantially affect the main experimental results.

In order to evaluate English language skills independent of their perceptual abilities, all subjects were given the written grammar portion of the *Oxford Placement Test I* (Allan, 1992). The two language groups did not differ significantly on this measure, $p > 0.05$. The average percentages of correct answers were 68% for Spanish speakers and 59% for Germans, indicating that the subjects predominantly had a lower-intermediate level of English competence (i.e., a functional, but not fluent, command of English).

2. Apparatus

The pre and post tests were conducted in quiet rooms, with stimuli played over headphones at a user-controlled comfortable level, and computers (PC and PDA) producing the stimuli and collecting responses. All training was conducted by subjects on their own; they borrowed PDAs, or the training software was installed on their own laptops. The training software created password-protected log files that the subjects could not access, so that we could verify that they completed all sessions.

All stimulus recordings were made in an anechoic chamber with 44 100 16-bit samples per second, and later downsampled to 11 025 samples per second.

3. Stimuli

a. Training

Recordings of English words were made from five speakers of British English, two male and three female. The words were groups of minimal pairs selected by dividing 14 British English vowels into four clusters: /ɛ/, /ɑ/, /a/, /ʌ/ (e.g., *pet, part, pat, putt*); /i/, /ɪ/, /aɪ/, /eɪ/ (e.g., *feel, fill, file, fail*); /u/, /əʊ/, /ɔ/ (e.g., *was, woes, wars*); and /u/, /aʊ/, /ɜ/ (e.g., *shoot, shout, shirt*). The clusters were selected by conducting a hierarchical cluster analysis on previous English vowel identification data by native Spanish and German speakers (Iverson and Evans, 2007); the first three clusters comprised vowels that were mutually confusable by many listeners, and the last cluster (i.e., /u/, /aʊ/, /ɜ/) was formed of remainders (i.e., vowels that were not strongly clustered with others). There were 10 sets of minimal pair words for each for the 4 clusters, for a total of 140 words. Each speaker recorded each word twice. All words were displayed to speakers one at a time in a random order during the recording session, to avoid list intonation.

b. Pre/post tests

Recordings of English /b/-V-/t/ words were made from two speakers of British English, one male and one female. Neither of these speakers and none of the words were used in the training corpus, such that all pre/post tests measured generalization to new stimuli. The speakers read the words *beat* /i/, *bit* /ɪ/, *bet* /ɛ/, *Burt* /ɜ/, *bat* /a/, *Bart* /ɑ/, *bot* /ɒ/, *but* /ʌ/, *bought* /ɔ/, *boot* /u/, *bait* /eɪ/, *bite* /aɪ/, *bout* /aʊ/ and *boat* /əʊ/; English vowels that would create non-words in the /b/-V-/t/ context (e.g., /ʊ/) were not included in the study. Four repetitions of these 14 English words were recorded for each talker, for a total of 112 stimuli.

A large set of synthesized stimuli from a previous study (Iverson and Evans, 2007) was used to map best exemplars. The stimuli were synthesized /b/-V-/t/ words embedded in a naturally spoken sentence frame (*Say_again*, spoken by a male speaker of British English), including the /b/ burst and the /t/ stop gap from the natural recording. The vowel stimuli were created using the cascade branch of a Klatt synthesizer (Klatt and Klatt, 1990). The synthesis parameters were chosen so that the synthesized vowel approximated the original vowel in the natural carrier sentence in terms of F0, amplitude contours, and spectrum. The stimuli primarily varied F1, F2, and duration, with some covaried variation in F3 (F3 was normally fixed to 2500 Hz, but was raised to be 200 Hz greater than F2 whenever F2 was greater than 2300 Hz). The F1 and F2 formant frequencies changed linearly from the beginning to the end of the vowel, and there were no additional consonantal formant transitions. F1 frequency was restricted so that it had a lower limit of 5 ERB (Glasberg and Moore, 1990) and an upper limit of 15 ERB. F2 frequency was restricted so that it had a lower limit of 10 ERB, was always at least 1 ERB higher than F1, and had an upper limit defined by the equation $F2=25-(F1-5)/2$. The stimuli were synthesized in advance with a 1 ERB spacing of the vowel space, and with 7 log-spaced levels of duration (54, 75, 104, 144, 200, 277, and 383 ms), for a total of 109 375 individual stimuli. The ERB and log-duration transforms allowed us to efficiently distribute the stimuli with regard to perception.

4. Procedure

a. Training

There were 5 sessions of high-variability phonetic training consisting of 225 trials of vowel identification with feedback, and an initial 14-trial practice session. The training sessions were run with no more than one session per day, and the entire course of training was completed over 1–2 weeks. The duration of each session was approximately 45 min. There was a different talker each session, and all subjects heard the talkers in the same order.

On each trial, subjects heard a stimulus word and clicked on three or four minimal-pair alternatives (depending on vowel; see stimulus description). For example, they could hear *slit* and be asked whether it sounded like *sleet, slit, slight*, or *slate*. The stimulus word was played before the response alternatives were shown, with the intent that the initial recognition of the word would be open set (e.g., not primed by the response alternatives), even though they gave a closed-set response. Every response word was accompanied by a more common word that had the same vowel (e.g., *seed, sit, night*, and *eight*) in case the response word was unfamiliar to the subjects. These example words were the same whenever that vowel appeared as a response, and the subjects were shown these example words during the initial instruction of the experiment.

If subjects gave a correct response, they saw "Yes!" on the computer screen accompanied by a cash register sound, then heard the word one more time. If subjects gave a wrong response, they saw "Wrong" on the computer screen accompanied by two tones with descending pitch, heard the correct word played, then heard a four-stimulus alternating series of the correct word, the incorrect response, the correct word, and the incorrect response. For example, if the stimulus word was *slit* and they clicked on the *sleet* response, they would hear an alternating series of *slit, sleet, slit*, and *sleet* so that they could learn the contrast between these two words.

Within each 225-trial training session, the first 70 trials were 5 repetitions of the 14 vowels in a random order, the next 85 trials were chosen adaptively based on the subject's errors, and the last 70 trials were 5 repetitions of the 14 vowels in a random order. This design ensured that all subjects were trained on all vowels at the beginning and end, while allowing some of the training to be customized to fit the needs of each individual subject. The adaptive trials were selected randomly, with the selection probability of an individual vowel being weighted by combining the proportions of misses and false alarms for that vowel. That is, the probability of a vowel being selected increased when it was identified incorrectly, or when that vowel was chosen incorrectly as a response when another stimulus had been played.

The stimulus words on each trial were chosen randomly for each vowel. That is, if the trial was intended to have an /i/ stimulus, the computer program randomly chose one of the ten minimal-pair stimulus words that had this vowel. This random selection was blocked, such that each of the ten minimal-pair word sets was used once before the list was recycled.

b. Pre/post tests

i. Vowel identification

Subjects heard natural recordings of English /b/-V-/t/ words and gave a closed-set identification response (all 14 words as response options). To give their response, they mouse clicked on a button which listed the stimulus word (e.g., *bot*) as well as a common English word that had the same vowel (e.g., *hot*). Prior to starting the experiment, they heard the speaker read a short story (i.e., *The North Wind and The Sun*) in order to familiarize them with the talker. They were shown the word response alternatives and were able to ask questions if they were unsure which vowels were indicated. The experiment was run twice (once for each talker), with four repetitions of each of the 14 vowels in a random order.

ii. L1 assimilation

Subjects heard natural recordings of English /b/-V-/t/ words and identified which of their own L1 vowels sounded closest to the vowel in the word that they heard. They were told that even though these were English vowels, they should

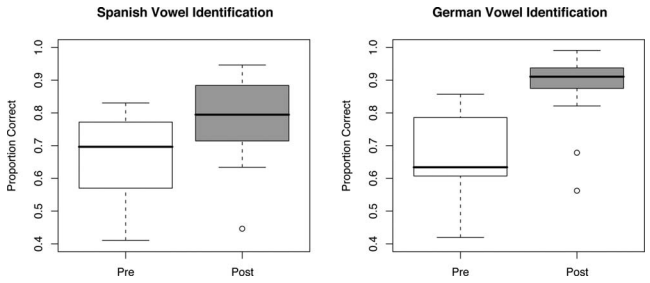


FIG. 1. Boxplots of the proportion correct vowel recognition in Experiment 1, pre and post training for Spanish and German speakers. Boxplots represent the quartile ranges of the scores, with outliers marked by circles.

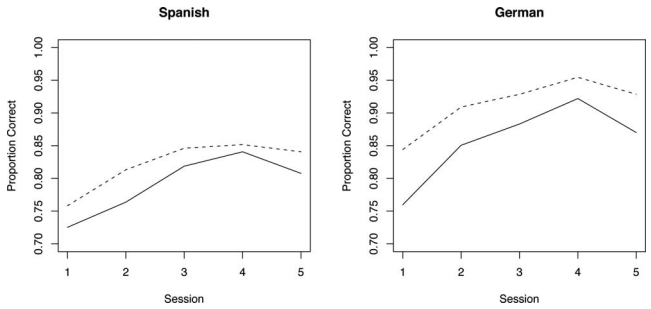


FIG. 2. Mean proportion correct at the start (solid line) and end (dashed line) of each training session in Experiment 1.

be classified as if they were listening to an L1 English speaker who was trying to speak their language. After each identification, they mouse clicked on a graphical continuum to rate whether this stimulus was *close* or *far away* from this L1 vowel category. The experiment was run twice (once for each talker), with two repetitions of each of the 14 vowels in a random order.

iii. Vowel-space mapping

On each trial, subjects saw an English /b/-V-/t/ word on the computer screen (e.g., *bot*), as well as a more common word that had the same vowel (e.g., *hot*), and heard a stimulus (synthesized /b/-V-/t/ embedded in a natural carrier sentence). They rated on a continuous scale how far away the /b/-V-/t/ that they heard was from being a good exemplar of the printed word. Their ratings were given by mouse clicking on a continuous bar presented on a computer screen.

A goodness optimization procedure (Evans and Iverson, 2004, 2007; Iverson and Evans, 2003, 2007; Iverson *et al.*, 2006) was used to iteratively change the stimuli that subjects heard on each trial, to search through the multidimensional stimulus space for good exemplars of each vowel. The full procedure will not be described here (see Iverson and Evans, 2007), but it involved simplifying the dimensionality of the search by finding best exemplars along straight-line paths that cut through the five-dimensional space, and efficiently choosing stimuli along each path so that they would be likely to be near to good exemplars (e.g., weighting the stimulus selection based on the subjects' previous responses). There were a total of seven search vectors and five trials per vector for each vowel. That is, subjects were able to find best exemplars after 35 trials, despite the large stimulus set (109 375 stimuli) and wide range of possible acoustic values available to subjects.

B. Results and discussion

1. Vowel identification

Figure 1 displays the vowel recognition accuracy for Spanish and German speakers before and after training. The subjects had been matched to minimize the pre-test differences between Spanish and German listeners, which is reflected in the similar pre-test boxplots. The post-test scores demonstrated improvement with training for both groups. However, the Spanish speakers improved relatively modestly (average 10 percentage points), while the German speakers improved twice as much (average 20 percentage points) and began to reach ceiling performance. These differences were confirmed with a repeated-measures analysis of variance (ANOVA) on arc-sin transformed scores. Specifically, there was a significant within-subject effect of pre/post, $F(1,24)=105.7$, $p<0.001$, demonstrating an overall improvement with training, and a significant interaction of pre/post and language group, $F(1,24)=14.7$, $p<0.001$, demonstrating that the two groups learned to different degrees; there was no main effect of language group, $p>0.05$. The results thus suggest that the relatively crowded vowel space of German speakers actually may have made vowel learning easier, rather than providing interference.

Hierarchical cluster analysis of the pre-test data revealed that Spanish subjects most frequently confused /i/-/ɪ/, /a/-/ʌ/-/ɑ/, and /ɒ/-/ɔ/; Germans most frequently confused /ɛ/-/a/-/ʌ/, and /ɑ/-/əʊ/-/ɒ/-/ɔ/-/aʊ/. After training, the two language groups improved both for these difficult clusters and for words that were not as strongly clustered. That is, there was a general improvement in vowel identification rather than a pattern of improvement that was markedly stronger or weaker for individual pairs. The improvement of Germans for /ɑ/-/əʊ/-/ɒ/-/ɔ/-/aʊ/ is notable because these words crossed the minimal-pair clusters used in the training. That is, Germans decreased their confusions for pairs like /ɒ/-/ʊ/ even though their forced-choice responses during training did not directly contrast these vowels.

In order to examine the improvement during the course of training, the accuracy for the first and last 28 trials (i.e., two repetitions of each vowel) was calculated for each training session (see Fig. 2). At the beginning of the first training session, the averages for Spanish and German listeners were relatively similar (0.73 and 0.76, respectively), but the Germans improved more on successive sessions. Compared across sessions, the results suggest that subjects may have approached asymptotic performance toward the end of the training session (i.e., the curve begins to flatten). However, the speakers occurred in the same order for each subject, so it cannot be determined, for example, whether the dip in performance on session 5 occurred because the subjects lost concentration or because the speaker that subjects heard in that session was difficult. Also, subjects continued to improve within each session even for sessions 4 and 5, suggesting that they were still learning in these sessions. To evaluate these observations statistically, a repeated-measures ANOVA analysis was conducted on arc-sin transformed scores. There was a significant main effect of session, $F(4,44)=11.8$, p

TABLE I. Assimilation of English vowels into L1 vowels, before and after training. The table lists the closest L1 vowel, with the percentage of trials chosen and its average perceived similarity (0 different to 1 same) listed in parentheses.

English vowel	Pre-test Spanish	Post-test Spanish	Pre-test German	Post-test German
i	i (100, 0.90)	i (100, 0.90)	i (72, 0.87)	i (95, 0.93)
ɪ	i (85, 0.88)	i (96, 0.88)	ɪ (84, 0.91)	ɪ (100, 0.94)
eɪ	e (100, 0.79)	e (100, 0.79)	ɛ: (40, 0.80)	ɛ: (50, 0.76)
aɪ	a (62, 0.81)	a (77, 0.79)	aɪ (88, 0.86)	aɪ (95, 0.92)
ɛ	e (96, 0.92)	e (96, 0.90)	ɛ (76, 0.88)	ɛ (82, 0.93)
a	a (92, 0.91)	a (100, 0.90)	a: (28, 0.84)	ɛ: (41, 0.77)
ʌ	a (81, 0.91)	a (88, 0.89)	a (68, 0.84)	a (82, 0.90)
ɑ	a (65, 0.87)	a (85, 0.83)	a: (68, 0.78)	a: (100, 0.88)
əʊ	o (100, 0.82)	o (100, 0.80)	o (72, 0.73)	o (77, 0.78)
ɒ	o (100, 0.91)	o (100, 0.88)	ɔ (96, 0.83)	ɔ (95, 0.91)
ɔ	o (100, 0.87)	o (100, 0.87)	o (88, 0.86)	o (95, 0.81)
aʊ	a (77, 0.77)	a (92, 0.78)	aʊ (96, 0.84)	aʊ (100, 0.92)
ɜ	e (46, 0.81)	e (62, 0.79)	ø (52, 0.70)	ø (55, 0.77)
u	u (100, 0.89)	u (96, 0.87)	u (52, 0.80)	u (68, 0.84)

<0.001, indicating that subjects improved as training progressed, and an interaction between session and language, $F(4,44)=2.8$, $p=0.036$, indicating that the Spanish and German speakers improved at different rates. There was a significant main effect of beginning/end, $F(1,11)=29.4$, $p<0.001$, indicating that listeners were better for the trials at the end of each session than they were at the start of each session, but there were no significant interactions with language or session, $p>0.05$. There was also a main effect of language, $F(1,148)=23.5$, $p<0.001$, indicating that Germans were more accurate overall than were Spanish speakers.

2. L1 assimilation

Table I lists how the stimuli assimilated into L1 categories before and after training. The patterns of assimilations only roughly correspond to the confusions made in identification. For example, the confusion of /ɒ/-/ɔ/ by Spanish speakers during the pre test makes sense because both sounded like /o/ in Spanish, but the assimilation ratings also predict that /əʊ/ should have been as frequently confused with these phonemes. The German /a/-/əʊ/-/ɒ/-/ɔ/-/aʊ/ confusions are poorly predicted by assimilation, because each of these vowels assimilated into different German categories. Part of this discrepancy could have been caused by the fact that these assimilation patterns were not very consistent (e.g., /ɑ/ assimilated most strongly into German /a:/, but did so only on 68% of the trials before training), indicating that there was some variability in how these stimuli were perceived.

The post-test results indicate that the assimilations may have changed somewhat after training, such that the closest L1 category was chosen more consistently. It is thus possible that the ability of these subjects to identify the English phonemes correctly was related to how consistently listeners assimilated them into their L1 categories. However, there is no established method for translating assimilation ratings into predictions of identification accuracy. If listeners were literally perceiving these stimuli in terms of their L1 categories, a plausible decision model for the identification task could be (1) listeners identify the English vowel in terms of an L1 category and (2) give an English response based on what English vowel usually sounds like that L1 category (e.g., a maximum-likelihood decision). For example, if a German listener was played English /ʌ/ and perceived it as closest to German /a/, it would be logical for them to give an answer of English /ʌ/ because that vowel most often sounds like German /a/. However, if a German listener was played English /ʌ/ and perceived it as closest to German /a:/, it would be logical for them to give a response of /ɑ/ (i.e., make an error) because /ɑ/ most often sounds like German /a:/. Although this model may be simplistic, one advantage of this decision model is that it can be easily translated into predictions from confusion matrix data. That is, each assimilation response on every trial can be translated into a maximum-likelihood English response (e.g., replacing all German /a/ assimilations with English /ʌ/), and then a predicted proportion of correct responses can be calculated.

This decision model was applied to the assimilation data for individual subjects, and compared to their actual identification accuracy scores. For German subjects before training, the predictions were very close. The average predicted accuracy (69%) was close to the obtained average accuracy in the identification experiment (68%), and there was no significant difference between these scores in a paired t-test, $p>0.05$. Moreover, individual differences in the predicted and obtained scores were significantly correlated, $r=0.69$, $p=0.009$. That is, individuals who were more accurate at identifying these phonemes also were more consistent at assimilating them to the closest German category. It is thus plausible that Germans before training were simply giving identification responses based on their assimilation into L1 categories.

However, this correspondence for Germans was weaker after training. The average predicted accuracy (77.6%) increased because of the increasing consistency of assimilation ratings, but the obtained identification accuracy increased more (87.8%). The two measures were significantly different, $t(10)=-2.72$, $p=0.021$, and the individual differences were no longer significantly correlated, $r=0.40$, $p>0.05$. It thus appears that training may have changed the extent to which the German subjects relied on L1 assimilation.

This assimilation model does not work at all for the Spanish subjects. Given that Spanish has only 5 vowels, the maximum performance that they could achieve would be 35.7% correct (i.e., 5 of the 14 English vowels correct). The decision model predicted that nearly all Spanish subjects would achieve this maximum before and after training, but their actual performance was much higher, averaging 66.8% before and 77.1% after training. The only way that Spanish subjects could plausibly have obtained these levels of performance was if they had been using additional categories or cues that they had learned for English, not just by assimilating these vowels into their Spanish categories.

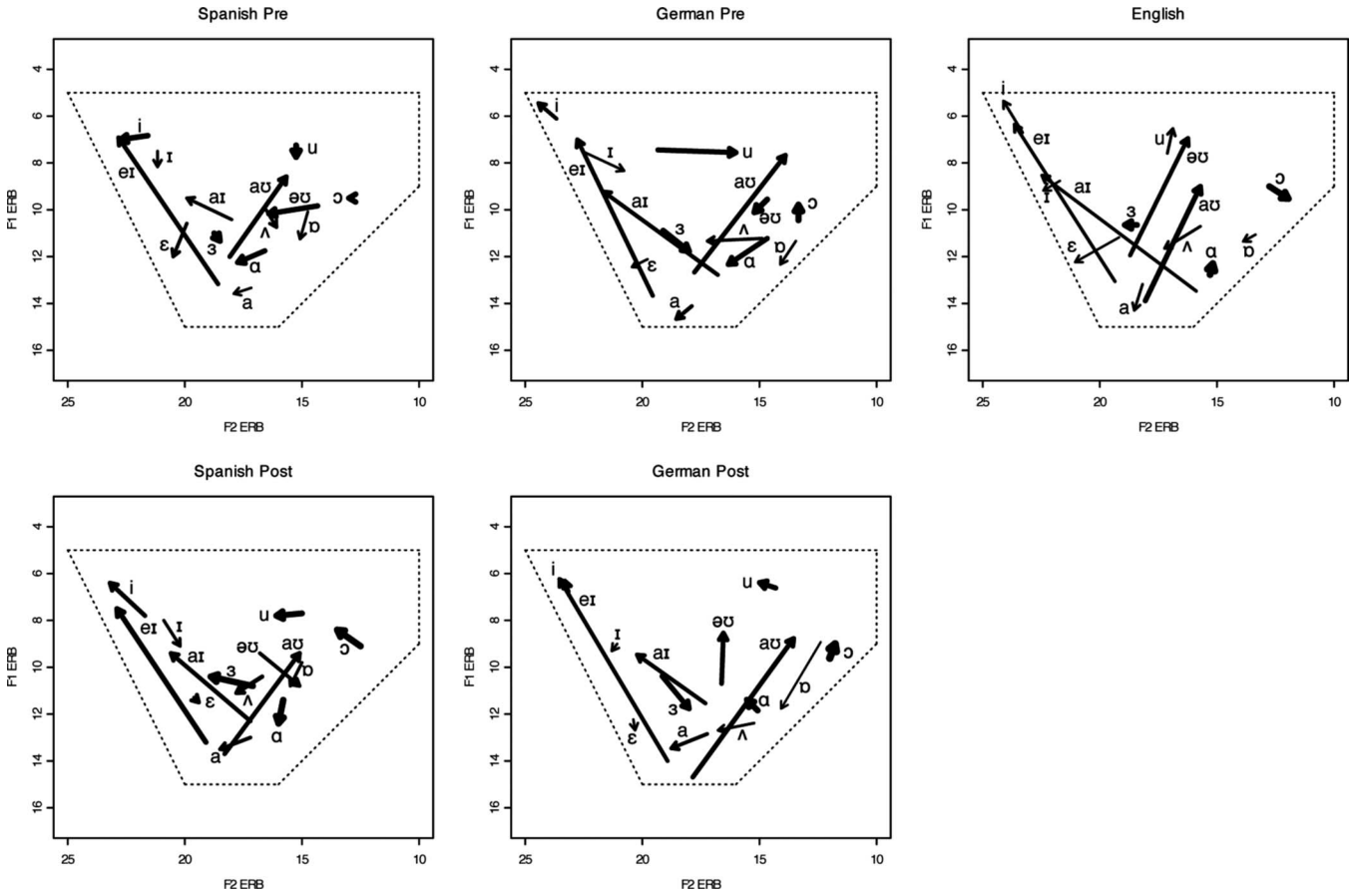


FIG. 3. Average best exemplar locations of English vowels in Experiment 1, for Spanish and German speakers pre and post training, and for L1 English speakers (from Iverson and Evans, 2007). Each vowel best exemplar is represented by a line from the starting to the ending first and second formant frequencies (i.e., indicating formant movement). Duration is indicated by the weight of the line, with thicker lines for longer vowels.

3. Vowel space mapping

Figure 3 displays the average best exemplars for Spanish and German speakers, as well as for English speakers from a previous study (Iverson and Evans, 2007). Although all language groups had generally similar vowel spaces (e.g., vowels in roughly the right locations, with appropriate degrees of formant movement), there were apparent differences. For example, Spanish and German speakers both had little contrast between /ʌ/ and /a/, and little format movement for /əʊ/, compared to English speakers.

In order to calculate how close the Spanish and German speakers were to the English averages before and after training, the distance was calculated between each individual's best exemplars and the average best exemplars for L1 English subjects (see Fig. 4). These distances were calculated separately for F1/F2 location, formant movement, and duration. The F1/F2 location accuracy was measured by averaging the beginning and ending frequency of each vowel for F1 and F2, giving a two-dimensional F1/F2 coordinate for that vowel with no formant movement. The Euclidean distance (i.e., root mean square) was then calculated between the F1/F2 locations of each individual's English best exemplars and the L1 English averages. Formant movement accuracy was measured by subtracting the F1/F2 location values above, so that each vowel had a vector representing the direction and magnitude of F1/F2 formant movement, with the center of each line passing through zero (i.e., normalizing the vowel's location in the vowel space). As above, Euclidean distances between these formant movement vectors were measured for each individual's vowels and the L1 English averages. Duration accuracy was quantified by calculating the average absolute-value difference between the durations of each individual's best exemplars and those of the L1 English averages.

Repeated-measures ANOVA analyses were conducted separately for each acoustic distance measure. For the measures of F1/F2 location and duration, there were no significant main effects or interactions of pre/post or language, $p > 0.05$; there was no evidence that listeners improved their vowel representations in these respects. There was a significant main effect of pre/post on formant movement, $F(1,16)=8.94$, $p=0.009$, although there was no main effect or interactions with language, $p>0.05$. Although the subjects thus improved in the formant movement of their best exemplars after training, this improvement was small (averaging 0.4 ERB), and the individual differences in this improvement were not significantly correlated with changes in identification accuracy, $r=0.30$, $p>0.05$.

Despite the fact that the accuracy of best exemplars did not improve much with training, identification accuracy was significantly correlated with individual differences in F1/F2 location accuracy, $r=-0.37$, $p=0.027$, and formant move-

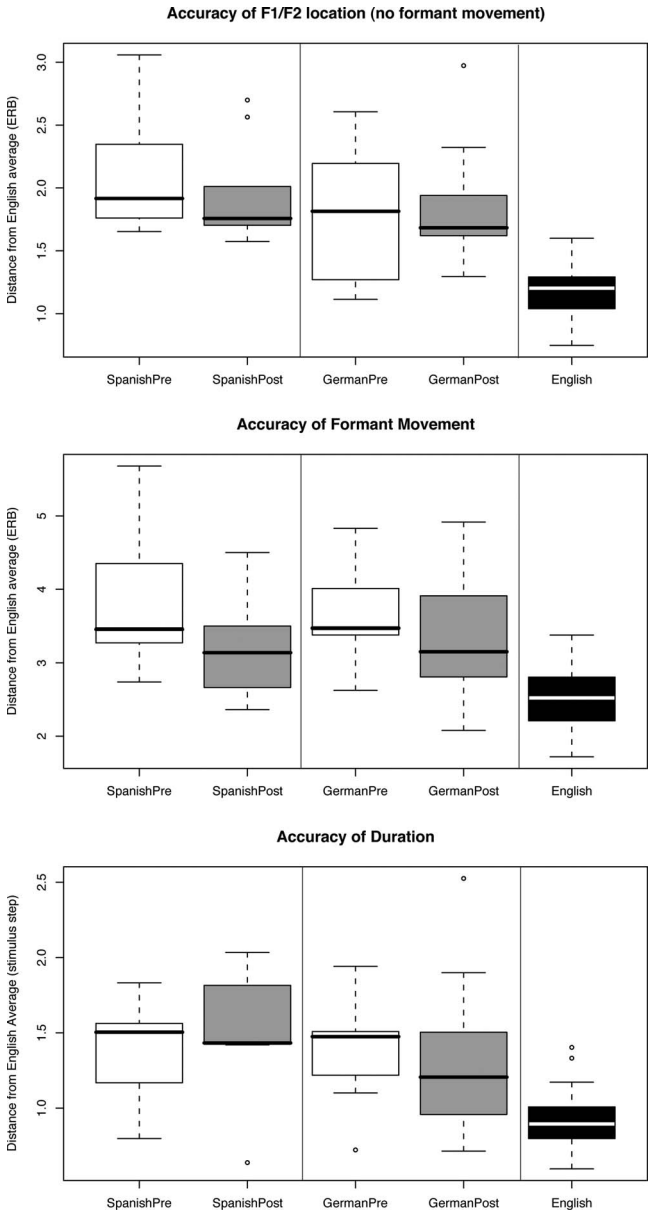


FIG. 4. Boxplots of the accuracy of each individual's best exemplar locations in Experiment 1 for Spanish and German speakers before and after training, and for L1 English speakers (from Iverson and Evans, 2007). The distances are listed separately for the accuracy of F1/F2 location, formant movement, and duration.

ment accuracy, $r=-0.41$, $p=0.014$, but not duration accuracy, $r=-0.05$, $p>0.05$. That is, individuals who were more accurate with their best exemplars in terms of F1/F2 location and formant movement tended to be better at identifying natural vowels, in accord with our previous study (Iverson and Evans, 2007).

III. EXPERIMENT 2: ADDITIONAL TRAINING AND RETENTION

There were two main unexpected results of Experiment 1. First, Spanish speakers learned less than the Germans. Their training results in each session began to flatten toward the later sessions, indicating that they may have been reaching an upper limit of learning, although the subjects continued to improve within each session. It is thus not clear whether the smaller improvements for Spanish speakers occurred because they reached an L1-related upper limit on their recognition accuracy (at least with this training method), or whether they simply learned slower than the Germans and could benefit from additional training. Second, the improvements in vowel identification accuracy for both groups were not accompanied by corresponding changes in best exemplars. One interpretation of this result is that the improvements in training may have been superficial. For example, the training could have served as a short term refresher of what they had learned previously about English vowels (e.g., in school), rather than producing long-term changes in their categorization processes. This hypothesis could help explain why Spanish subjects learned less; they were already using English while living in London, so they could have benefited less from "refreshing" than did Germans, who had not been using English recently.

Experiment 2 was conducted to investigate these possibilities. The Spanish speakers in Experiment 1 were contacted an average of 4 months later and asked to complete an additional ten sessions of training. The German subjects were contacted 1 year later and were asked to complete only the identification task. The timing of these tests was based on circumstance, rather than being a planned aspect of the experimental design, and we tested whatever subjects in Experiment 1 were willing to participate (i.e., all subjects rather than only the matched subgroups). However, this design still met the goals of evaluating whether both groups of listeners were able to retain their improvement over long intervals, and whether Spanish listeners could further improve with additional training.

A. Method

1. Subjects

The subjects were 9 Spanish and 11 German speakers. They were recruited from the group of subjects who participated in Experiment 1, including 2 Spanish subjects who had participated in a pilot version of the training protocol which was similar to the final version but had a more random selection of stimuli.

The Spanish subjects were tested 2–6 months (mean 4 months) after they completed Experiment 1. They completed all pre/post tests and a ten-session course of training, as described below. The German subjects were tested 1 year later, and completed only the vowel identification test.

2. Procedure

The pre/post tests were the same as in Experiment 1. The training protocol was expanded to ten sessions by recording stimuli from five additional speakers (three male and two female), but was the same in all other respects.

B. Results and discussion

1. Vowel identification

The identification results (Fig. 5) suggest that both Spanish and German speakers retained their training improvements in Experiment 1. For Spanish speakers, the mean accuracy in the Experiment 1 post test (0.76) was similar to

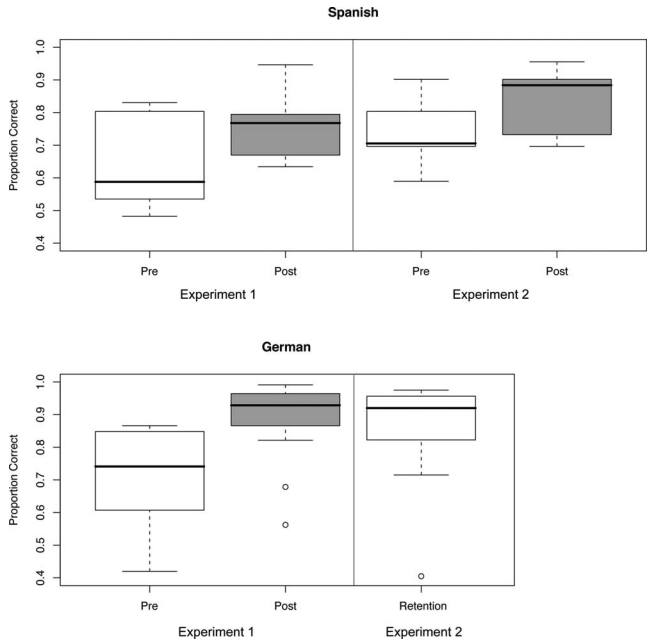


FIG. 5. Boxplots of the proportion correct vowel recognition in Experiments 1 and 2. The Experiment 1 data are displayed only for the subset of subjects who participated in both experiments.

that obtained 4–5 months later (0.73), and there was no significant difference between these scores, $t(8)=-0.92$, $p>0.05$. However, the conclusion that there was retention is tempered by the fact that there was also no significant difference between their accuracy in the Experiment 1 pre test (mean=0.64) and the Experiment 2 pre test, $t(8)=-2.04$, $p=0.076$. The retention results for German subjects are clearer; their accuracy 1 year after training (mean=0.85) was not significantly different from their Experiment 1 post-test accuracy (mean=0.88), $p>0.05$, and was significantly greater than their Experiment 1 pre-test accuracy (mean=0.71), $t(10)=6.50$, $p<0.001$.

Spanish speakers significantly improved from the additional training in Experiment 2, raising their accuracy from a mean of 0.73 at the pre test to 0.84 at the post test, $t(8)=-5.24$, $p<0.001$. This result demonstrates that they had not reached a ceiling in performance in Experiment 1, and were able to achieve similar levels of performance as German subjects had in Experiment 1 after they were given additional training. Spanish speakers may thus take longer to learn English vowels than do German speakers, but still have the same basic capacity to learn.

2. Vowel space mapping

Despite the fact that Spanish speakers had more training, Fig. 6 demonstrates that the accuracy of their vowel space mapping did not improve. There were no significant differences between pre and post tests in terms of how close their best exemplars were to the average best exemplars for English, in terms of F1/F2 location, formant movement, or duration, $p>0.05$. The results were thus much the same as in Experiment 1.

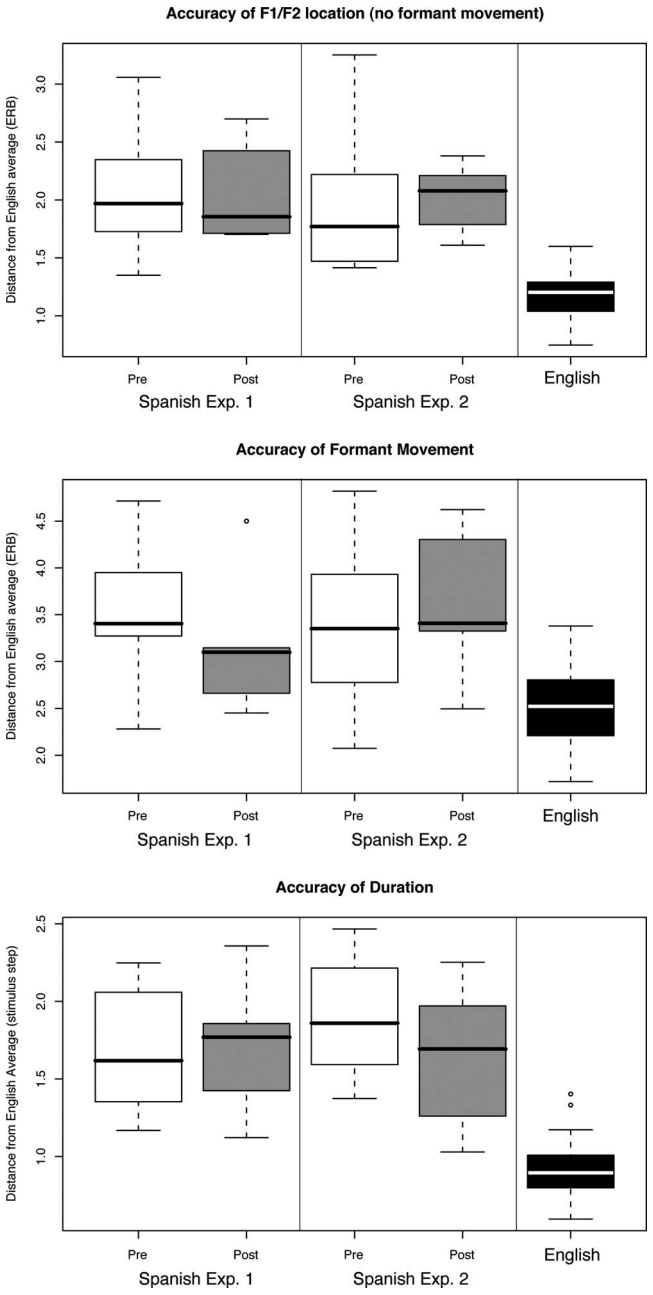


FIG. 6. Boxplots of the accuracy of each individual's best exemplar locations in Experiments 1 and 2, for Spanish before and after training, and for L1 English speakers (from Iverson and Evans, 2007). The distances are listed separately for the accuracy of F1/F2 location, formant movement, and duration.

IV. GENERAL DISCUSSION

The results demonstrate that German and Spanish speakers learn at different rates given auditory training; German speakers improved twice as much, on average, as Spanish speakers after five sessions of training, although Spanish speakers attained similar levels of performance after completing an additional ten training sessions. One of our working hypotheses based on SLM (Flege, 1995) had been that learning would be more difficult for the Germans; their crowded L1 vowel space should have left less room for learning new vowel categories. However, the German speakers learned more easily than did the Spanish speakers, sup-

porting our previous evidence of learning by individuals with large L1 systems (Iverson and Evans, 2007). Although this result may thus appear to be contrary to SLM, there was little evidence that German subjects actually formed new categories during the experiment (i.e., no improvement in the accuracy of best exemplars); without category learning there would be no expectation for there to be L1 category interference.

How did identification accuracy improve without corresponding changes in best exemplars? One could initially question the methods used in this study, raising doubts about whether subjective notions of best exemplars are related to the processes underlying vowel recognition. However, the validity of the best exemplar task is well supported by previous work; the accuracy of best exemplars has been significantly correlated with individual differences in the accuracy of English vowel recognition by L2 speakers (i.e., Experiment 1, and more strongly with the wider range of subjects tested in Iverson and Evans, 2007), vowel production and speech-in-noise recognition by English speakers with different accents (Evans and Iverson, 2004, 2007), vowel recognition by cochlear implant users (Iverson *et al.*, 2006), and with the identification of English /r/-/l/ by Japanese speakers (Hattori and Iverson, 2009).

The task used to map best exemplars has relatively low processing demands, in that listeners are able to replay the stimulus and only have to listen for one word at a time. In the present study, it assessed whether listeners had an accurate subjective notion of what vowels sound good in English. However, it is one thing to "know" what vowels sound good in English and another to put this knowledge in practice. For example, the recognition of vowels in natural speech involves more variability in the stimulus (e.g., related to talker differences or phonetic environment), more possible responses, and only a single time to hear the stimulus; listeners must rapidly encode the phonetic information that is relevant to the L2 categorization. Likewise, the perception and production of real-world speech (e.g., in a conversation) involves yet a higher processing load (i.e., less opportunity to focus only on the phonetic content of speech when listeners need to concentrate on meaning). Our hypothesis based on the present results is that auditory training improves the ability of subjects to apply their existing category knowledge to natural variable speech (i.e., both L1 and L2 category knowledge), without changing this knowledge itself. That is, auditory training makes the categorization process more efficient and automatic in a way that is long lasting, but does not generally change the representation of the categories (e.g., use of cues).

The original high-variability phonetic training papers hypothesized that training was successful because it changed attentional weights for phonetic dimensions, causing listeners to attend more to aspects of the stimuli that can distinguish categories over a range of talkers and phonetic environments and less to irrelevant variability (e.g., Lively *et al.*, 1993; Logan *et al.*, 1991). High variability training was thought to generalize better than training with single talkers or small stimulus sets, because, in part, the higher variability taught listeners which cues were most robust. However, there has been no evidence that this reweighting of cues occurs due to this kind of training. For example, Iverson *et al.* (2005) trained Japanese adults on English /r/-/l/ using stimuli that had been signal processed in different conditions to alter the cues that they heard (e.g., enhancing differences along F3 or manipulating the variability of secondary dimensions such as F2 and transition duration). Japanese adults learned in all conditions, but did not change the cues that they used to match the acoustic manipulations of the conditions. Instead, they appeared to become more consistent at labeling stimuli as English /l/ when the stimuli became acoustically similar to a Japanese flap (i.e., short closure, short transition; e.g., see Hattori and Iverson, 2009), even when this decision conflicted with the acoustic information that they had received in training. That is, listeners became better at applying their existing L1 knowledge to this task, without necessarily improving in the cues that they used for /r/ and /l/. Similarly, Heeren and Schouten (2008) found that Dutch speakers can be trained to identify the Finnish /t/-/t:/ contrast, but do this without obtaining the peak in discrimination sensitivity at the category boundary that is typical of native Finnish speakers for these stimuli, suggesting that auditory training may not produce pervasive changes in the way that individuals perceive these phonetic contrasts.

There is some evidence that laboratory based training can alter the use of acoustic cues, but this work has all involved restricted stimulus sets or entirely novel categories. For example, Francis and colleagues found in several studies that listeners change perceptual weightings for stimuli following identification training, but this has occurred for relatively homogenous sets of synthesized speech (i.e., no simulated talker differences; Francis *et al.*, 2000, 2007) and for a novel Korean stop-consonant contrast produced by a single talker (Francis and Nusbaum, 2002). Likewise, Holt and Lotto (2006) trained listeners to learn a novel category for non-speech sounds (frequency-modulated tones) and found that listeners developed cue weightings that were affected by the distribution of the training stimuli (i.e., relying less on dimensions that were more acoustically variable). It is thus clear that individuals can, in principle, learn the cues of new categories and re-weight the cues of existing categories when given laboratory-based auditory training. Although this work may still be useful for understanding auditory categorization, there is little evidence that changes in cue weightings are the primary learning mechanism in typical high-variability phonetic training studies, where individuals have previous experience with the categories and are trained on natural speech from multiple talkers.

High-variability phonetic training may thus be more effective than training with less variable stimulus sets, because the stimulus variability trains the process of applying categories to real speech. That is, listeners prior to auditory training typically have some idea about what the various categories sound like, and training makes them more efficient at applying this knowledge to situations in which they do not know exactly what they will hear from trial to trial, and when there is phonetic variability that is irrelevant to the categorization judgement. Whereas training with less variable stimulus sets (e.g., Lively *et al.*, 1993) simply does not train this kind of

ability; listeners are able to learn to categorize small stimulus sets without improving in a more general ability that can be applied to other stimuli. If high-variability training improves the application of category knowledge to variable speech more than it alters cue weightings, it may not be necessary for the variability in the training set to be fully natural in order for performance to improve (e.g., Iverson *et al.*, 2005).

In the present study, German listeners may thus have learned faster than Spanish listeners because they had more pre-training category knowledge to apply to the task. For example, the Germans had more L1 vowel categories, and the analyses of assimilation ratings of Experiment 1 indicated that part of their improvement with English vowels may have occurred because they became more consistent at assimilating them into German categories. Both German and Spanish speakers likely had some English category knowledge prior to training, which would also be expected to become applied more automatically after training. One could suspect that the English language experience of Spanish speakers could have made them benefit less from training, because they may have already been using their category knowledge more efficiently due to their daily experience with using English. However, there was no correlation between their amount of time living in English-speaking countries and their improvement due to training in Experiment 1, and Experiment 2 demonstrated that they had a capacity to learn if given more time (i.e., they had not reached a ceiling for this type of learning). Moreover, we have recently examined the role of experience more directly (comparing the English vowel perception of L1 French speakers who were living in France or the UK), and found that individuals who used English every day improved from training as much as individuals who had little spoken English experience (Iverson and Preece-Pinet, 2008).

Although high-variability phonetic training may primarily improve recognition by making existing categorization processes more efficient, it is clear that long-term exposure to phonetic categories (e.g., when living in an English-speaking environment) creates changes to the cues that individuals use. For example, differences between subjects in their best exemplars of English vowels are predictive of vowel identification ability (Iverson and Evans, 2007), and L1 Spanish speakers can weight English vowel duration differently depending on how duration is used by the English accent that they primarily hear (Escudero and Boersma, 2004). Therefore, although high-variability phonetic training improves identification performance, it may not provide a full simulation for the kinds of changes in phonetic perception that occur during longer term L2 language learning.

ACKNOWLEDGMENTS

The authors are grateful to Anke Sennema for subject recruitment and hosting the data collection at University of Potsdam (Germany). This research was funded by Grant No. RES-000-23-0838 from the UK Economic and Social Research Council.

Allan, D. (**1992**). *Oxford Placement Tests 1*, Oxford University Press, Oxford, UK.

Best, C. T. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Language Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore) pp. 171–204.

Best, C. T., McRoberts, G. W., and Goodell, E. (**2001**). "American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology," J. Acoust. Soc. Am. **1097**, 775–794.

Bohn, O.-S. (**1995**). "Cross-language speech perception in adults: First language transfer doesn't tell it all," *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 279–304.

Bohn, O.-S., and Flege, J. E. (**1990**). "Interlingual identification and the role of foreign language experience in L2 vowel perception," Appl. Psycholinguist. **11**, 303–328.

Delattre, P. (**1965**). *Comparing the Phonetic Features of English, French, German, and Spanish* (Harrap & Co., London).

Escudero, P., and Boersma, P. (**2004**). "Bridging the gap between L2 speech perception research and phonological theory," Stud. Second Lang. Acquis. **26**, 551–585.

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.

Evans, B. G., and Iverson, P. (**2007**). "Plasticity in vowel perception and production: A study of accent in young adults," J. Acoust. Soc. Am. **121**, 3814–3826.

Flege, J. E. (**1989**). "Differences in inventory size affect the location but not the precision of tongue positioning in vowel production," Lang Speech **32**, 123–147.

Flege, J. E. (**1995**). "Second language speech learning: Theory, findings, and problems," *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 233–277.

Flege, J. E. (**2003**). "Assessing constraints on second-language segmental production and perception," *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, edited by A. Meyer and N. Schiller (Mouton de Gruyter, Berlin).

Flege, J. E., Bohn, O.-S., and Jang, S. (**1997**). "The effect of experience on nonnative subjects' production and perception of English vowels," J. Phonetics **25**, 437–470.

Flege, J. E., Schirru, C., and MacKay, I. R. A. (**2003**). "Interaction between the native and second language phonetic subsystems," Speech Commun. **40**, 467–491.

Francis, A. L., and Nusbaum, H. C. (**2002**). "Selective attention and the acquisition of new phonetic categories," J. Exp. Psychol. Hum. Percept. Perform. **28**, 349–366.

Francis, A. L., Baldwin, K., and Nusbaum, H. C. (**2000**). "Effects of training on attention to acoustic cues," Percept. Psychophys. **62**, 1668–1680.

Francis, A. L., Nusbaum, H. C., and Fenn, K. (**2007**). "Effects of training on the acoustic phonetic representation of synthetic speech," J. Speech Lang. Hear. Res. **50**, 1445–1465.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Gottfried, T., and Beddor, P. S. (**1988**). "Perception of spectral and temporal information in French vowels," Lang Speech **31**, 57–75.

Hattori, K., and Iverson, P. (**2009**). "English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy," J. Acoust. Soc. Am. **125**, 469–479.

Heeren, W. F. L., and Schouten, M. E. H. (**2008**). "Perceptual development of phoneme contrasts: How sensitivity changes along acoustic dimensions that contrast phoneme categories," J. Acoust. Soc. Am. **124**, 2291–2302.

Hirata, Y., Whitehurst, E., and Cullings, E. (**2007**). "Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates," J. Acoust. Soc. Am. **121**, 3837–3845.

Holt, L. L., and Lotto, A. J. (**2006**). "Cue weighting in auditory categorization: Implications for first and second language acquisition," J. Acoust. Soc. Am. **119**, 3059–3071.

Iverson, P., and Evans, B. G. (**2003**). "A goodness optimization method for investigating phonetic categorization," in the Proceedings of the 15th International Conference of Phonetic Sciences, Barcelona, Spain.

Iverson, P., and Evans, B. G. (**2007**). "Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration," J. Acoust. Soc. Am. **122**, 2842–2854.

Iverson, P., and Preece-Pinet, M. (**2008**). "Training English vowels for French speakers with varying English experience," J. Acoust. Soc. Am. **123**, 3734.

Iverson, P., Hazan, V., and Bannister, K. (**2005**). "Phonetic training with

acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults," J. Acoust. Soc. Am. **118**, 3267–3278.
Iverson, P., Smith, C. A., and Evans, B. G. (**2006**). "Vowel recognition via cochlear implants and noise vocoders: Effects of formant movement," J. Acoust. Soc. Am. **120**, 3998–4006.
Kingston, J. (**2003**). "Learning foreign vowels," Lang Speech **46**, 295–349.
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.
Lambacher, S., Martens, W., Kakehi, K., Marasinghe, C., and Molholt, G. (**2005**). "The effects of identification training on the identification and production of American English vowels by native speakers of Japanese," Appl. Psycholinguist. **26**, 227–247.
Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). "Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories," J. Acoust. Soc. Am. **94**, 1242–1255.
Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). "Training Japanese listeners to identify English /r/ and /l/: A first report," J. Acoust. Soc. Am. **89**, 874–886.
Meunier, C., Frenck-Mestre, C., Lelekov-Boissard, T., and Le Besnerais, M. (**2003**). "Production and perception of foreign vowels: does the density of the system play a role?" in the Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain.
McAllister, R., Flege, J. E., and Piske, T. (**2002**). "The influence of the L1 on the acquisition of Swedish vowel quantity by native speakers of Spanish, English and Estonian," J. Phonetics **30**, 229–258.
Morrison, G. (**2002**). "Perception of English /i/ and /ɪ/ by Japanese and Spanish listeners: Longitudinal results," in *Proceedings of the North West Linguistics Conference 2002*, edited by G. S. Morrison and L. Zsoldes (Simon Fraser University Linguistics Graduate Student Association, Burnaby, BC, Canada), pp. 29–48.
Nishi, K., and Kewley-Port, D. (**2007**). "Training Japanese listeners to perceive American English vowels: Influence of training sets," J. Speech Lang. Hear. Res. **50**, 1496–1509.
Stockwell, R. P., and Bowen, J. D. (**1965**). *The Sounds of English and Spanish* (University of Chicago Press, Chicago).
Strange, W., Bohn, O.-S., Nishi, K., and Trent, S. (**2005**). "Contextual variation in the acoustic and perceptual variation of North German and American English vowels," J. Acoust. Soc. Am. **118**, 1751–1762.
Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., and Munhall, K. G. (**2008**). "Training English listeners to perceive phonemic length contrasts in Japanese," J. Acoust. Soc. Am. **123**, 397–413.
Trubetzkoy, N. S. (**1969**). *Principles of Phonology*, translated by C. A. M. Baltaxe (University of California Press, Berkeley).

Experimental jet velocity and edge tone investigations on a foot model of an organ pipe

Hubert J. Außerlechner[a] and Thomas Trommer
Chair of Building Physics, University of Stuttgart, Pfaffenwaldring 7, 70569 Stuttgart and Fraunhofer Institute for Building Physics, Nobelstraße 12, 70569 Stuttgart, Germany

Judit Angster
Fraunhofer Institute for Building Physics, Nobelstraße 12, 70569 Stuttgart, Germany

András Miklós
Steinbeis Transfer Center Applied Acoustics, Landauer Straße 24, 70499 Stuttgart, Germany

(Received 27 October 2008; revised 28 May 2009; accepted 31 May 2009)

In order to investigate the physical processes involved in the build-up of the sound signal in a labial organ pipe a pipe foot model has been developed. The main important parameters, such as positions of the lower and upper lips, the wind pressure in the pipe foot, and the width of the flue, can be adjusted by means of this model. Moreover, different types of languids and pipe bodies (resonators) can be attached to the model. For the reason of corresponding to a real metal organ pipe these parts of the model are made of a typical alloy used in organ building. The reproducibility of measurements is provided by the micrometer screws applied for the adjustments. Flow and edge tone measurements are carried out with the help of this model. Velocity measurements with different flue widths show that the exit velocity of the jet corresponds to the Bernoulli-velocity and is asymmetrically contracted. At larger distances (>5 mm) the velocity distribution can be described by a Gauss-function having linearly increasing width. A mathematical relation of the centerline velocity as a function of the cut-up height L is found. The results of edge tone measurements show differences between previous studies and the present one. No frequency stages (and hysteresis phenomena) are found within the investigated pressure and cut-up range; the frequency modes of the edge tone coexist. The measured frequencies can be described by theoretical models.
 [DOI: 10.1121/1.3158935]

PACS number(s): 43.75.Np, 43.28.Ra, 43.28.Tc [NHF] Pages: 878–886

I. INTRODUCTION

When a jet leaves a slit and impinges against an edge a tone is generated. This phenomenon, called edge tone, may influence the attack transient and the stationary sound of a flue organ pipe. The edge tone frequency depends mainly on the exit velocity of the jet, the slit-to-edge distance, and the flue width.

Detailed experimental investigation of sound generation in open flue organ pipes is one of the main subjects of a research project running presently at the group of musical acoustics of the Fraunhofer Institute for Building Physics. Since the first acoustic signal that appears in the pipe is an edge tone, it is reasonable to investigate the properties of edge tone and its effect on the build-up of the pipe tone in detail. For that reason a pipe foot model has been developed for studying jet and edge tone properties. The construction of the model allows the attachment of a pipe resonator in order to investigate the build-up of the pipe sound. The results of the experimental research will be published in several papers. The present work, which is limited to jet and edge tone investigations, is the first paper of the planned publications.

Some authors say that the pitch of the edge tone changes gradually for a while (first stage), then it jumps suddenly to another pitch corresponding to the next stage. Furthermore a hysteresis is observed by increasing and decreasing the parameter.[1–4] This means that at a certain value of the increasing parameter the frequency suddenly jumps from the lower stage to the higher one and it jumps back to the lower stage at a lower value of that parameter when decreasing the parameter. Other authors do not find such phenomena. Jones[5] defined two types of edge tones. The first type occurs at low velocities and exhibits frequency jumps, while the second one appears at higher velocities without the jumps. It is also observed that the frequency jumps depend on the used velocity profile at the flue exit. Krothapalli and Horne[6,7] found out that top hat profiles at the flue permit the coexistence of several discrete tones, whereas Poiseuille profiles (comply with a parabolic profile) do not.

Many studies have been published on the edge tone. Brown[1] first presented an empirical formula of the edge tone frequency for a 1 mm flue width:

$$f = 0.466j(U_0 - 40)(L^{-1} - 0.07), \tag{1}$$

which is expressed in terms of the mean velocity U_0 (in cm/s) of the emitting jet, the slit-to-edge distance L (in cm), and j, where j defines different values for the frequency stages. Here the frequency decreases with L to the power of

[a]Author to whom correspondence should be addresses. Electronic mail: hubert.ausserlechner@ibp.fhg.de

−1 (exponent $n=-1$), whereas Jones[5] used the exponents −1.00, −1.14, −1.22, and −1.42 for the first four frequency stages. In terms of the Strouhal number, $\mathrm{St}=fd/U_0$, where d is the width of the flue, Holger[8] and Crighton[9] found that the measured frequency fits better with $n=3/2$:

$$fd/U_0 = \alpha(d/L)^{3/2}. \tag{2}$$

The parameter α is dependent on the frequency stage. A recent consideration of Paál and Vaik[10] again shows good results with $n=-1$. A good review of the edge tone and the theories of its generation are given by Karamcheti *et al.*[11] In Ref. 12 differences between the sounding frequencies of the edge tone and the organ pipe are investigated. As well the phase delay between acoustic pressure and the crossing of the jet at the edge of the upper lip is measured.

A nonlinear edge tone model including the interactions of the jet with the upper lip is given by Howe,[13] Verge *et al.*,[14,15] and Fabre *et al.*[16] Due to the separation of the jet flow, and for higher Reynolds numbers, the acoustic flow at the edge of the upper lip vortex shedding occurs. This vortex shedding affects the amplitude and the generation of high frequencies of an edge tone.

The mentioned edge tone models have worked well for basic edge tone investigations; however, they are not suitable for modeling the sound generation in organ pipes. The geometries of the models (pipe foot, flue channel, edges, etc.) and the ranges of the adjusted parameters (pressure in the pipe foot, position of the edge, flue width, etc.) are quite different from the parameter ranges used by organ builders. In Ref. 17 different voicing steps at a metal organ pipe are described. It is shown that small changes in the mouth area, the region of the emitted jet, gradually change the amplitude and the frequency of the pipe sound.

Pitsch *et al.*[18,19] performed several acoustical measurements with an edge tone model on the basis of a diapason 4′ C^0 pipe and flow visualization experiments in a two-dimensional pipe foot model in a flat-water channel. The main disadvantage of their model is the low reproducibility of the adjustable parameters.

This paper describes the development of a pipe foot model with accurately adjustable parameters in dimensions which are used in the organ building. The dimensions of the model are consistent with the Töpfer scale (normal scale). Also the flow and acoustic measurements performed by means of this model are done with parameters of real organ pipes. In the organ building ratios $L/d>10$ are used. Further the geometry of the flue channel of a labial pipe is always asymmetrical. The pipe foot is conical and has a very short flue channel build by the edge of the languid and lower lip (cf. Fig. 1). Another important part is the shape of the edge of the upper lip. Symmetrical sharpened edges (wedges) are never used in organ building. Organ builders say that pipes do not sound "good" with such edges. That is the reason for the use of symmetrical (rectangular) or slightly rounded upper lips. And last but not least the pressure in the pipe foot is in the range of $p_{\mathrm{pf}}=300,\ldots,1200$ Pa. These values are consistent with exit velocities of $U_0\approx 22,\ldots,44$ m/s. Because of higher Reynolds numbers and the geometry of the model the jet can be considered as turbulent rather than laminar.

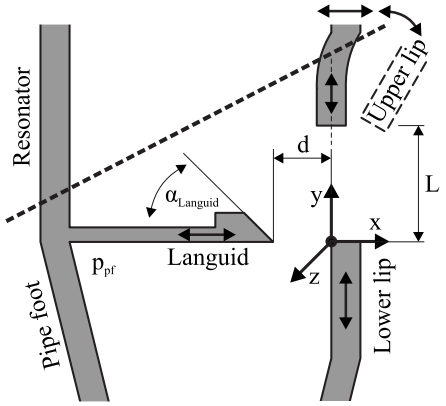


FIG. 1. Sketch of the mouth area of a labial organ pipe, description and adjustment possibilities (arrows) of the parameters (cut-up L, flue width d, languid angle $\alpha_{\mathrm{Languid}}$, and the pressure in the pipe foot p_{pf}), and definition of the origin of the coordinate system.

The behavior of the jet (exit angle and profile) and the acoustic behavior at different parameter settings can probably help understand the sound production of real organ pipes. If the effects of changing the parameters are physically comprehended, the voicing of an organ pipe can be simplified.

II. PIPE FOOT MODEL

An organ pipe, a sketch of the mouth area is shown in Fig. 1, consists of a so called pipe foot (the lower part), the resonator (the long tube) attached to the pipe foot, and the languid (plate inside the pipe). The opening at the front side is called mouth or cut-up L and is defined by the lower and upper lips (labium). The slit between the languid and the lower lip is called flue with the flue width d.

If the resonator is omitted the remaining part may be regarded as a pipe foot model (this corresponds to the right side of the dashed line). Most of the past researches have dealt with completely different edge tone models which are not similar to real labial organ pipes. For example, they offer a symmetrical alignment of the (in most cases symmetrical) languid, a very long air channel from which the jet leaves, untypical geometries (flue width, cut-up height, and languid), too high or two low air pressures, or different kinds of materials but not organ metal. Hence a pipe foot model with following demands has been developed (on the basis of the similarity of the model to a real organ pipe the model is called pipe foot model instead of edge tone model).

(1) The dimensions should be consistent with a diapason 4′ C^0 organ pipe (diameter of the resonator 84 mm).
(2) To the similarity to real organ pipes the most common alloy used by organ builders (60% lead and 40% tin) should be used for the relevant parts (languid, lower and upper lips, and resonator).
(3) Languid and resonators of different kinds can be attached to the model. This means resonators with different lengths and/or diameters (up to 100 mm) and languids with different languid angles and profiles.
(4) The most important feature: high reproducibility of the measurements. The positions of the upper and the lower lip and of the languid can be adjusted in micrometer

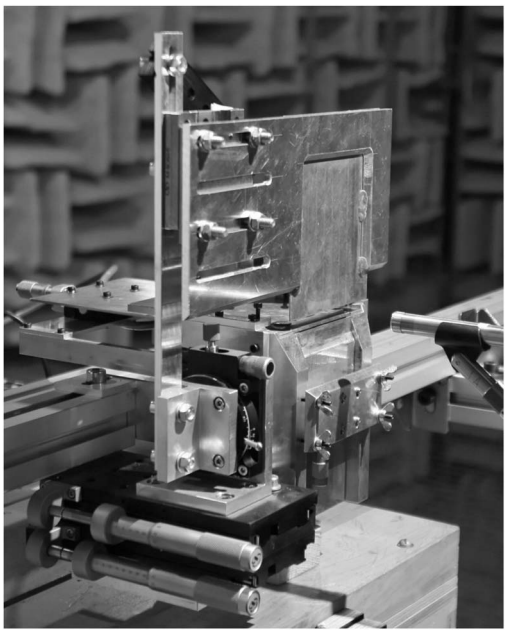


FIG. 2. Pipe foot model with the adjusting devices. At the bottom the two linear tables and the rotation table of the upper lip, at the front side the micrometer screw of the lower lip, and at the back side the plate of the languid and its linear table. On this plate different kinds of languids can be mounted.

steps. The smallest step size of the micrometer screws (four linear translation stages and one rotation table, see Fig. 2) is 1 μm.

(5) The angle of the upper lip can be changed.
(6) Possibility of variable pressure adjustments (pressure measurement inside the pipe foot) is given.

Figure 2 shows a photo of the developed pipe foot model. The whole pipe foot is made of aluminum except for the languid. To allow for a better rigidity the base plate of the languid is made of steel. In the organ building the height of the lowest pipe foot is defined by the "standardized" height of the rack pin and the pipe rack with a value of about 180–185 mm. This length is held constant until resonator diameters of 52 mm. Because of stability problems (long and heavy resonators) and design (prospect pipe or form of the lower lip) the pipe foot length is increased for larger diameters. Here none of these problems are relevant and so the total height of the model is fixed at 150 mm and the flue breadth at 60 mm. The body is conical with the diameters 30 and 100 mm.

For a better stability the adjusting devices are mounted by angular brackets on the pipe foot model. The lower lip is attached to a linear translation stage with an adjustment range of 10 mm and adapted into a channel to allow a perpendicular alignment of the edge corresponding to the languid. Two linear translation stages allow a travel range of nearly 100 mm for the upper lip. Additionally, a rotation table is added to allow different angle settings. In order to avoid deformation, the upper lip is attached to an aluminum plate. As mentioned above the base plate of the languid is made out of steel of 2 mm thickness. Different kinds of organ metal languid profiles (languid edge angles of 45° and 60° and some special languid bevels) can be mounted on the base plate without deformation. The whole pipe foot model is installed on an aluminum rack. The possibility of attaching a resonator onto the model is not given, however shown in Fig. 2 and not discussed in this paper.

In Fig. 1 all possible adjustments of upper and lower lips and the languid are shown with arrows. The pressure in the pipe foot p_{pf} is measured by means of a 0.4 mm hole at the opposite side of the lower lip very close to the bottom of the languid plate. The origin of the coordinate system is defined as the intersection of the plane of the bottom of the languid, the plane of the inner surface of the lower lip, and the vertical plane of the centerline of the upper lip (see Fig. 1). If all coordinates are equal to zero, theoretically no jet can be emitted.

III. EXPERIMENTAL SETUP

A blower, a diaphragm bellows, a wind trunk, and a small slider chest represent the organ instrument. The pipe foot model is placed on the slider chest. The pressure in the pipe foot and thus the flue velocity can be adjusted either by changing the weight on the diaphragm bellows or by regulating the fan speed of the blower. To minimize the noise and fluctuations of the blower different kinds of mufflers combined with flow stabilization tools are inserted between the blower and the diaphragm bellows. All measurements have been carried out in an anechoic room.

Sound pressure is measured by a B&K type 4165 condenser microphone placed horizontally at a distance of 50 mm in front of the lower lip (underside of the microphone and upper side of the lower lip are on the same height) to avoid recording of unwanted blowing noise. The microphone is calibrated using a B&K type 4230 sound level calibrator (94 dB, 1000 Hz).

Another very important measurement device is the anemometer for velocity measurements. A SVMtec hot wire anemometer [3D-Flow-4CTA (Ref. 20)] with a 10 μm sensor is used, so measurements very close to the flue can be done. Due to the upper frequency limit of 3.5 kHz the jet fluctuations can also be measured. An extra feature of this system is the auralization of the fluctuating parts (extracted by the software) of the measured velocity. This feature is very useful, for example, for the localization of the boundary layers of the emitted jet.

The pressure inside the pipe foot can be measured with a Honeywell 163PC01D36 pressure sensor. In addition to this sensor a second one is used to detect the pressure in the slider chest. For averaging and checking the results an HP 35670A dynamic signal analyzer is used.

IV. EXPERIMENTAL RESULTS

A. Jet velocity measurements

To determine the velocity profile of the undisturbed jet velocity measurements have been performed by using the pipe foot model without the upper lip (cf. Fig. 1). The flue width d is changed in 0.4 mm steps from 0.5 to 2.1 mm corresponding to the values used in real organ pipes. All velocity measurements are carried out with a constant pressure of

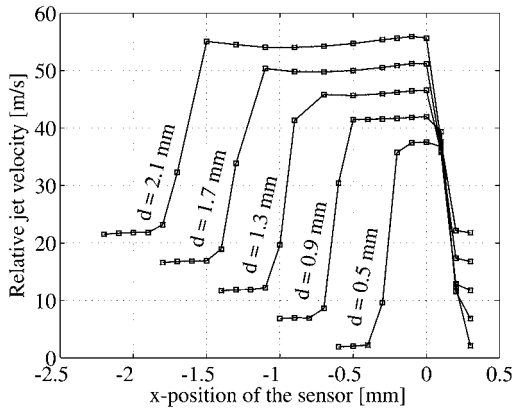


FIG. 3. Jet velocity profiles at the flue at different flue widths with the 45° languid. For better illustration all profiles except the profile at d=0.5 mm are shifted: d=0.9 mm: +5 (m/s), d=1.3 mm: +10 (m/s), d=1.7 mm: +15 (m/s), and d=2.1 mm: +20 (m/s).

700 Pa (±2 Pa) in the pipe foot. For a single velocity measurement the signal of the hot wire anemometer is recorded for 3 s at a sampling rate of 22.05 kHz. The mean value and the standard deviation [root mean square deviation (RMSD)] of the recorded data are calculated. The mean value is regarded then as the measured velocity, while the RMSD value as the measured velocity fluctuation.

By means of the 10 μm sensor of the anemometer, it is possible to measure the velocity as close as y=0.5 mm above the flue. Velocity profiles of the 45° languid measured at five different flue widths (0.5, 0.9, 1.3, 1.7, and 2.1 mm) are shown in Fig. 3.

The solid lines with squares represent the velocity profiles measured with the 45° languid. (In all diagrams the symbols, such as squares, circles, and triangles, represent the measured data points.) Due to the very short flue channel (it is defined by the lower lip and the edge of the languid) a top hat velocity profile can be expected. The ideal top hat velocity profile is defined as

$$U(x) = \begin{cases} 0, & x > 0 \\ U_0, & -d < x < 0 \\ 0, & x < -d. \end{cases} \quad (3)$$

In reality the jet has boundary layers of finite thickness which cause the so-called Nolle velocity profile.[21]

It can be seen that the velocity profiles at the flue are similar to Nolle profiles and are shifted to the right. The almost constant parts of the profiles start around x=0, i.e., over the inner edge of the lower lip (see Fig. 1) and end at a negative x-value, which is larger than the corresponding negative d-value. It means that the width of the velocity profile at a distance of y=0.5 mm above the flue is smaller than the flue width, i.e., the jet is asymmetrically contracted to about 50%–70% of the flue width. The contraction coefficient ε (the ratio of the cross sectional area of the jet and the flue area) is smaller for narrower flue; its value can be estimated as 0.5, 0.56, 0.62, 0.65, and 0.71 for flue widths 0.5, 0.9, 1.3, 1.7, and 2.1 mm, respectively.

The measured velocities (34–37 m/s) correspond quite well to the free outflow velocity at 700 Pa pressure difference. At the range of constant velocity of the RMSD, the measure of the velocity fluctuations is also constant around the small value of 0.6 m/s (<2% of the maximal velocity). Within the boundary layers the RMSD value rapidly increases up to about 6 m/s.

TABLE I. Average velocities (U_0) for the 45° and the 60° languid at ranges of constant RMSD and different flue widths d.

d (mm)	$U_{045°}$ (m/s)	$U_{060°}$ (m/s)
0.5	36.88	37.13
0.9	36.69	36.83
1.3	36.13	36.16
1.7	35.47	35.76
2.1	34.93	35.16

The measured velocity profiles of the two used languids (languid edge angles of 45° and 60°) are similar. They differ only close to the languid edge. The edge represents the separation point of the flow. This means that the air stream detaches itself from the surface of the languid and vortex shedding occurs. The more acute-angled the edge of the languid is the less is the possibility of the jet to reattach at the languid.

In Table I the average of the velocities measured at y =0.5 mm and within the x-range defined by a constant RMSD is shown. With Bernoulli's equation

$$U_{0B} = (2p_{\mathrm{pf}}/\rho_0)^{1/2}, \quad (4)$$

where ρ_0 is the air density, the maximal velocity U_{0B} at the flue can be calculated as 34.1 m/s. The average velocities U_0 at y=0.5 mm in Table I are slightly greater than U_{0B} and fall linearly with increasing flue width. At higher velocities air cannot make sharp turns easily. So the jet separates at the edge of the languid and the jet is constricted (vena contracta effect:[22] the velocity increases at this region). This effect is larger for a smaller flue; thus it can explain both observed phenomena; the higher velocities at y=0.5 mm as U_{0B} and the slight decrease in U_0 with increasing flue width. However, the total change of about −2 m/s is only about 5%–6% of the average velocity.

Since both x- and y-components of the velocity contribute to the velocity signal of the hot wire sensor, higher velocity values close to the jet boundary may be resulted by inflow at both boundaries to the jet. In addition to the vena contracta effect the jet draws air into itself (entrainment effect[23]). This inflow effect can be seen in Fig. 3 (best at the profiles of flue width d=2.1 mm). At both ends of the quasi-horizontal part of the velocity profile the velocity is higher than within this range.

No significant conclusion can be drawn from the velocity differences between the two languid angles because the values lie within the error range of the anemometer (within these velocity ranges the error is less than 0.15 m/s).

Because of the x-symmetry of the pipe foot model the same velocity profile along the flue is assumed. Two measurements at the half (z=30 mm) and the quarter (z =15 mm) of the flue length have been carried out and show that the assumption of symmetry is true. At the range of

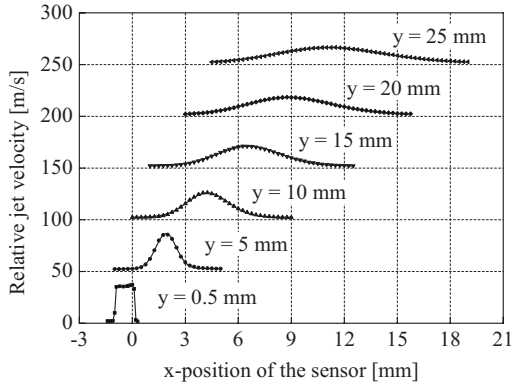


FIG. 4. Measured and fitted jet velocity profiles at different distances to the flue (flue width $d=1.3$ mm and 45° languid). For better illustration all profiles except the profile at $y=0.5$ mm are shifted: $y=5$ mm: +50 (m/s), $y=10$ mm: +100 (m/s), $y=15$ mm: +150 (m/s), $y=20$ mm: +200 (m/s), and $y=25$ mm: +250 (m/s).

constant velocity no discrepancies occur between the two profiles; at the boundary layers of the jet the differences become larger. This may be an effect of surface imperfections of the languid and the lower lip. Minor pressure fluctuations also may have an effect on the velocity. The velocity profiles at both ends of the flue breadth are not measured ($z=0$ mm and $z=60$ mm).

For larger distance y to the flue the jet spreads outward in the x-direction, the entrainment of the surrounding fluid increases, and the centerline velocity $U(x_c,y)$ decreases so that the total momentum remains constant[23] (see discussion below). This spread can be described by different profiles.

Bickley profile:[24,25]

$$U(x,y) = U(x_c,y)\mathrm{sech}^2(x/b(y)), \tag{5}$$

modified Bickley profile or Nolle profile:[21]

$$U(x,y) = U(x_c,y)\mathrm{sech}^2((x/b(y))^n), \tag{6}$$

and the Gaussian distribution:

$$U(x,y) = u_0 + \left(\frac{A(y)}{w(y)(\pi/2)^{1/2}}\right)\exp\left(-2\left(\frac{x-x_c(y)}{w(y)}\right)^2\right). \tag{7}$$

Here $b(y)$ represents the half width parameter of the jet and n a positive integer (for better fits decimal numbers are used). The Gaussian parameters are the offset u_0, the $1/e^2$ width $w(y)$, the area $A(y)$, and the center $x_c(y)$ of the distribution. The best fits have been reported by Gaussian curve fits[26] followed by Nolle's modified Bickley profile and last by the Bickley profile. So in the following fits are done with the Gaussian distribution.

Velocity profiles are measured at six different heights above the flue (at y-values of 0.5, 5, 10, 15, 20, and 25 mm) for each flue width within an x-range of +0.3 to $-(d+0.1)$ mm (where d is the actual flue width). The data points in Fig. 4 show the velocity profiles at different distances ($y=0\cdots25$ mm in 5 mm steps) with flue width $d=1.3$ mm and for a languid angle of 45°; the curves of the 60° languid are similar. The position of the lower lip is $y=0$ for all measurements. The solid lines in Fig. 4 are the fitted Gaussian functions. The excellent agreement shows that the spreading of the jet can be described by a Gaussian velocity distribution. The offset u_0 represents the entrainment of air to the jet from both sides; it is not necessary for characterizing the velocity distribution.

TABLE II. Calculated jet angles θ_{jet} and the width of the fitted Gaussian profiles w for the 45° languid and the 60° languid at flue width of $d=1.3$ mm and different y-distances (position of the lower lip is $y=0$ mm).

	45° languid		60° languid	
y (mm)	θ_{jet} (deg)	w (mm)	θ_{jet} (deg)	w (mm)
5	27.14	1.16	27.01	1.20
10	26.20	2.34	25.99	2.36
15	25.82	3.55	25.54	3.60
20	25.67	4.75	25.36	4.79
25	25.55	5.92	25.24	5.91

At the flue and at a very small distance the centerline velocity is nearly constant. Then it decreases with increasing distance y. A similar result is empirically found by Yoshikawa and Arimoto.[27] Their findings for the "organ 1" foot model agree well with the results in this paper but only for velocity profiles at longer distances to the flue. As shown in Fig. 7 of Ref. 27, the Nolle profile still exists at a distance of 8 mm from the flue.

With these data a mathematical relation of the velocity reduction can be found:

$$U(x_c,y) = C_u y^{-1/2}, \tag{8}$$

where $C_{u45}=81.87$ mm$^{1/2}$ m s^{-1} and $C_{u60}=80.31$ mm$^{1/2}$ m s^{-1} for the two languid angles, respectively. Furthermore it is possible to calculate the jet exit angle θ_{jet}. In Table II the results of this calculation and the w-parameters ($1/e^2$ width of the Gaussian profile) are listed. The exit angles decrease slightly with increasing distance.

Closer investigation of the exit angle data shows that the centerline of the jet over 5 mm distance can be regarded as a straight line shifted slightly to the positive x-direction. For the 45° languid this straight line starts at $x_{45}=0.215$ mm with an angle of $\theta_{45}=(25.16\pm0.02)°$, while for the 60° languid the corresponding data are $x_{60}=0.239$ mm and $\theta_{60}=(24.81\pm0.04)°$. The exit angle of the 60° languid is slightly smaller than that of the 45° languid (see discussion of the velocity profiles at the flue). This effect is also found by Yoshikawa and Arimoto.[27]

In the organ building the direction of the jet is mainly modified by changing the position of the languid regarding the position of the lower lip, but different languid angles are also used to control the jet exit angle.[28] The results of the velocity measurements show that the languid angle insignificantly affects the jet exit angle.

The $1/e^2$ width $w(y)$ of the Gaussian velocity profile increases linearly with y. Linear functions of the form

$$w(y) = ay \tag{9}$$

are fitted to the values of Table II. Both fits have coefficients of determination R^2 of nearly 1 which would be consistent

with the original function (R^2_{45}=0.99989, a_{45}=0.2367 ± 0.0005; R^2_{60}=0.99991, a_{60}=0.2394 ± 0.0005). The slope of the line is slightly larger for the 60° languid. Due to the asymmetry of the model and Reynolds numbers between 1133 (d=0.5 mm) and 4760 (d=2.1 mm) within the range of transitional flow,[29] the jet should probably be considered as turbulent rather than laminar. As for the laminar flow, the y-components of the momentum of the fluid must be constant[22,23]

$$M = \int_S \rho U(x,y)^2 dS = \text{const.} \tag{10}$$

By substituting the Gaussian velocity distribution from Eq. (7) into Eq. (10), the momentum M can be calculated for flue breadth b as

$$M = \rho b A(y)^2 / w(y) / \pi^{1/2} = \text{const.} \tag{11}$$

That is, the quantity $A(y)^2/w(y)$ should be constant for a given flue geometry. This requirement for the fit parameters is fulfilled for all investigated jets, i.e., the momentum is conserved by the jet emitted from the flue.

Introducing the centerline velocity $U(x_c,y)$ instead of $A(y)$ and taking into account that the constant value of M equals the momentum at the flue,

$$M = \rho b \varepsilon d U_{0B}^2, \tag{12}$$

the centerline velocity can be calculated as

$$U(x_c,y) = 2^{1/2} \pi^{-1/4} U_{0B} (\varepsilon d / w(y))^{1/2}. \tag{13}$$

The contraction coefficient of the issuing jet is ε=0.62.

B. Edge tone measurements and frequency analysis

For edge tone measurements the upper lip is added to the model. The upper lip is parallel positioned to the z-axis. The x- and y-positions and the angle of the upper lip to the y-axis can be adjusted by micrometer screws. Since the pitch and loudness of the edge tone depend significantly on the position of the upper lip related to the velocity profile, the angle and the x-position are adjusted so that the amplitude of the first edge tone mode is the highest. The variable parameters are the flue width, the slit-to-edge distance L (cut-up), and the jet exit velocity. The velocity is changed by the pressure in the pipe foot.

The outflow direction and velocity profile of the jet did not change when the upper lip was attached. However, slight changes have been observed in the case of a strong edge tone. As the maximal velocity decreased, the Gaussian profile became broader and the standard deviation of the velocity increased. The changes became smaller when the loudness of the edge tone was reduced by shifting the upper lip away from the optimal x-position. Such behavior can be expected when the jet oscillates in the transverse direction.

If we listen to an edge tone we notice no stable tone. It is an "active" tone with fluctuating pitch and loudness and of course some flow noise can also be heard. Spectra of typical edge tones are shown in Fig. 5 for different pressures in the pipe foot (100, 200, and 700 Pa). Several peaks can be seen

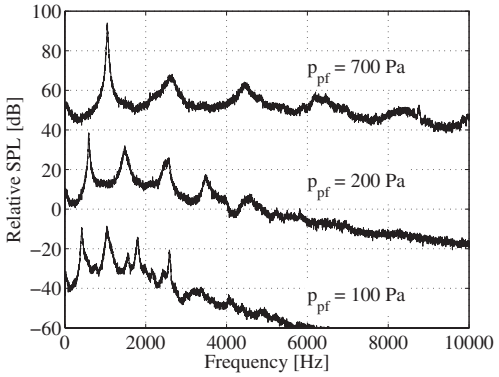


FIG. 5. Typical edge tone spectra with the 45° languid, the cut-up L =10 mm, the flue width d=1.3 mm, and with different pressures in the pipe foot (p_{pf}=100 Pa: absolute SPL-curve has been shifted −80 dB; p_{pf} =200 Pa: absolute SPL-curve has been shifted −40 dB and p_{pf}=700 Pa). The upper lip is positioned at the position at which the amplitude of the first mode is the highest.

in the spectrum. Normally one peak is much narrower and higher than the others. The frequency and amplitude of this peak determine the pitch and loudness of the listened edge tone. A gradual change in either the wind pressure or the cut-up resulted in gradual shifts of the frequencies and amplitudes of the peaks. Moreover, additional peaks may appear, or already existing peaks may disappear. Up to seven or eight peaks have been found in the measured edge tone spectra.

In the whole frequency range (0···22.05 kHz) of the spectra of Fig. 5, p_{pf}=700 Pa, six peaks can be observed. These peaks correspond to different sinuous oscillation modes[21,30] of the planar jet. The presence of several sinuous modes in the edge tone spectrum means that the jet can oscillate at more frequencies at the same time; i.e., the edge tone modes coexist.

The modes are often called "stages" in the literature because sudden changes in the pitch of the edge tone have been observed in several earlier experiments.[1,2,5,31] Since in the presented work all measured edge tone spectra show the coexistence of several peaks, and only gradual changes in the peak frequencies and amplitudes are observed, it is more appropriate to use the expression "mode" instead of stage because the latter implies the existence of discrete, separated sounding domains of the edge tone.

1. Varying the pressure (velocity)

The pressure in the pipe foot is changed in 100 Pa steps from 100 to 1000 Pa at different flue widths (d =0.5–2.1 mm) and languids. For low pressures the spectral peaks are very close together with a small bandwidth (cf. Fig. 5, p_{pf}=100 Pa). If the pressure increases the distance between the bandwidth of the peaks becomes wider (cf. Fig. 5, p_{pf}=200 Pa).

During the gradual change in the mentioned parameters the dominant mode may also change. For example, when the pressure in the pipe foot decreased from 200 to 100 Pa, the highest amplitude is shifted from the first mode to the second one in Fig. 5. When the edge tone is listened to by a person, the gradual increase in the amplitude of the second mode

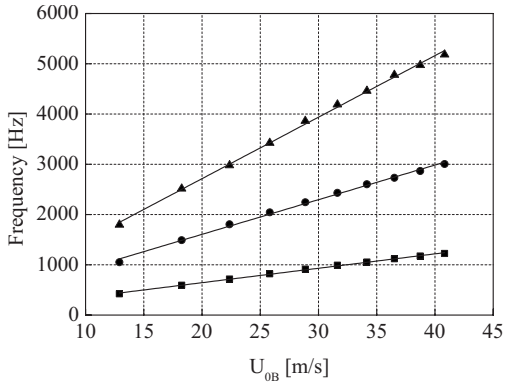


FIG. 6. Measured frequency of the first three modes (mode I: squares, mode II: circles, and mode III: triangles) of the edge tone as a function of the calculated mean exit velocity U_{0B} of the jet (cut-up $L=10$ mm and flue width $d=1.3$ mm). Results of a linear curve fit (solid) are plotted for comparison of linearity.

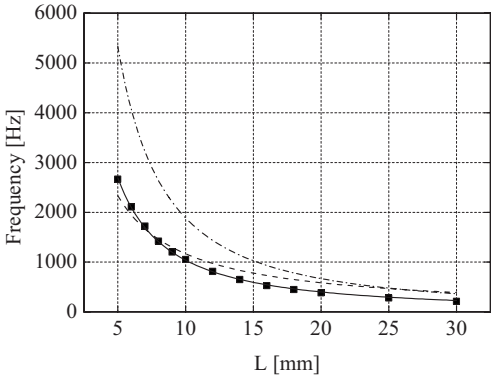


FIG. 7. Cut-up dependency of the frequency mode I: flue width $d=1.3$ mm and pressure in the pipe foot $p_{pf}=700$ Pa. Measured data points are plotted as symbols. The solid lines denote the results by the modified formula [Eq. (16)], the dashed lines the results of Bamberger and Páal [Eq. (15)], and the dashed dotted lines those of Holger [Eq. (14)].

may be resulted in a sudden jump of the perceived pitch from the original one (determined by the first mode) to a higher pitch (determined by the second mode). This subjective effect may explain the earlier observations about the existence of separated stages of edge tone.

In several earlier published papers sudden frequency jumps and hysteresis between two stages have been observed when the driving pressure of the jet was changed. However, no such effect has been observed in our experiments within the investigated pressure range of 100–1000 Pa. The attention of this work is turned to the development of a pipe foot model with respect to real organ pipes. Therefore other materials, geometries, and parameter settings are used as in the earlier studies, which dealt with the basic physical processes.

The edge tone frequency calculations are performed by three different theoretical models: the model of Holger:[8]

$$f_N = 0.925(N+\alpha_N)^{3/2}U_0 d^{1/2}/L^{3/2}, \tag{14}$$

the very common form of Bamberger *et al.*[32] and Paál and Vaik:[10]

$$f = C_d U_0/L, \tag{15}$$

and a modified version of Eq. (14) investigated by the authors and first proposed in this publication:

$$f_N = C_{d,N}U_0 d^{1/2}/L^{n_{d,N}/2}. \tag{16}$$

In Eqs. (14)–(16) N denotes the mode number and α_N is a numerical constant. Its values for the first three modes are $\alpha_1=0.4$, $\alpha_2=0.35$, and $\alpha_3=0.5$, C_d, $C_{d,N}$, and $n_{d,N}$ are fit constants, and U_0 is the mean exit velocity of the jet. For constant flue width d and cut-up L all the above three equations are linear functions of the velocity[33] U_0. However, the slopes of the linear functions are different. Measured (symbols) edge tone frequencies of the first three modes are plotted in Fig. 6 for $d=1.3$ mm and $L=10$ mm as functions of the calculated exit velocity U_{0B} [Eq. (4)]. The linear dependency of the frequencies on U_{0B} is confirmed by a linear fitted curve (solid). The higher the mode is, the steeper the slope of the curve is.

The frequency ratios $f_2/f_1=2.5$ and $f_3/f_1=4.25$ are constant within the whole pressure range, but they decrease with increasing flue width. At time no appreciable relation of the amplitude changes in different modes and in different flue widths has been found. A first approximation can be given: the measured data show that the amplitudes of the first mode increase with increasing pressure while the amplitudes of the other modes increase until 400 Pa then keep nearly constant or they fall slightly.

2. Varying the slit-to-edge distance

In this section the dependency of the edge tone frequency on the cut-up L is discussed. Measurements are done with a constant pressure of 700 Pa in the pipe foot, and again with different values of the flue width d (0.5–2.1 mm) and with the two languids (45° and 60°). Measured and calculated edge tone frequencies are plotted in Fig. 7 for the first mode as a function of the cut-up L. An increase in the cut-up resulted in the decrease in the frequencies and the amplitudes. Again no frequency jumps occur. The number of modes became continuously smaller. The functions and calculations of the higher modes are similar to those of mode one, but shifted.

In the figure three frequency curves, calculated by Eqs. (14)–(16), are plotted too. With the formula and parameter settings of Holger the deviation is much too high. The calculations with L^{-1} differ as well as from the measured data. Parameter settings for Eq. (16) according to Table III give the best fit results for all measured flue width and cut-up settings.

Another approach of calculating the edge tone frequency deals with the assumption of a feedback loop.[2,34] The disturbances of the jet need a travel time τ_1 from the flue to the

TABLE III. Fit parameters (constant $C_{1.3}$ and exponent $n_{1.3}$) of the modified Eq. (16) of Holger for the first three frequency modes at flue width of $d=1.3$ mm.

	$C_{1.3}$ (−)	$n_{1.3}$ (−)
Mode I	600.59	2.76
Mode II	1166.65	2.57
Mode III	1908.57	2.53

TABLE IV. Fit parameters of Eq. (18) for the first three frequency modes at flue width of $d=1.3$ mm

	ΔL (m)	C_u ($m^{3/2}/s$)
Mode I	6.3267×10^{-4}	0.788 58
Mode II	11.5×10^{-4}	2.157 67
Mode III	11.3×10^{-4}	3.943 56

upper lip. The acoustic signal that generated a little distance downstream the tip of the edge of the upper lip[11,35,36] has a travel time $\tau_2=(L+\Delta L)/c_0$ back to the flue, where it produces a disturbance on the jet. Here ΔL and c_0 are the distance of the sound source from the upper labium and the speed of sound, respectively. Thus, a feedback loop with an acoustical feedback exists. This feedback can maintain an edge tone oscillation with a frequency of $f=1/(\tau_1+\tau_2)$. It is assumed that disturbances propagate with about half of the flow velocity[30] of the jet. According to Eq. (8) the centerline velocity of the jet is proportional to $y^{-1/2}$; therefore the propagation velocity can be written as $C_u y^{-1/2}$, where C_u is a constant. Thus τ_1 can be calculated by integrating the propagation velocity from the flue to $L+\Delta L$:

$$\tau_1 = \int_0^{L+\Delta L} \frac{y^{1/2}dy}{C_u} = 2(L+\Delta L)^{3/2}/3C_u. \tag{17}$$

The frequency of the edge tone can be given as

$$f = \left(\frac{L+\Delta L}{c_0} + \frac{2(L+\Delta L)^{3/2}}{3C_u}\right)^{-1}. \tag{18}$$

Regarding C_u and ΔL as fit parameters (see Table IV) Eq. (18) fits excellently to the measured edge tone frequencies. This result also explains why neither the Holger model nor the Paál–Bamberger model could properly describe the dependence of the edge tone frequency on the cut-up distance. It follows from Eq. (18) that if f is approximated by a power function, its exponent must be larger than $-3/2$ and smaller than -1. This condition is fulfilled by the modified Holger-equation applied in the present paper.

By the edge tone measurements the y-position (cut-up L) of the upper lip is adjusted first, then the upper lip is moved along the x-axis until the position is found, where the amplitude (measured and averaged by the HP analyzer) of the first edge tone mode is maximal. From the adjusted x- and y-positions an angle can be calculated and compared to the measured jet exit angle given in Table II. The calculated angle values correlate well with the data of Table II, showing that the maximal edge tone amplitude is produced at the maximum of the velocity profile.

V. CONCLUSION

For an experimental study on the edge tone a pipe foot model with adjustable parameters has been developed. The applied geometry and materials are based on real organ pipes. By means of micrometer resolution linear and rotation tables, the pipe foot model allows reproducible measurements and a high accuracy of the adjustments.

Velocity measurements with different flue widths show that the exit velocity of the jet corresponds to the Bernoulli-velocity, and it is asymmetrically contracted. At larger distances (>5 mm) the velocity distribution can be described by a Gauss-function having linearly increasing width. A mathematical relation of the centerline velocity as a function of the distance to the flue y is found and used for further frequency calculations. By substituting the found velocity distribution into the momentum conservation equation its validity for the jet has been confirmed. The results of the presented work allow the determination of flow direction and velocity profile from the flue geometry and the pressure in pipe foot. This information can help the organ builder by the voicing adjustment of labial organ pipes.

The results of edge tone measurements already show differences between previous studies and the present one. No frequency stages (and hysteresis phenomena) are found within the investigated pressure and cut-up range; the frequency modes of the edge tone coexist. The measured frequencies can be described by the modified Holger model or the model of Eq. (18). Significant differences in the flue geometry and exit velocity range compared to the flue geometries and velocity ranges of previous investigations may explain the different results. The acoustic measurements also confirm the jet exit angle calculated from the velocity measurements. If the upper lip is positioned at the center of the velocity profile the edge tone amplitude is maximal.

In this paper the basic flow and edge tone phenomena of the pipe foot model have been investigated. The traditional knowledge of organ builders about the sensitivity of the labium on small adjustments is fully supported by the results. However, the effects of the different parameters on the jet and edge tone have been measured by modern instruments not only by listening the pipe sound. The expressions found for jet velocity distribution and edge tone frequencies may facilitate the very complex work of voicing adjustment of organ pipes.

ACKNOWLEDGMENTS

The authors acknowledge the financial support of the Deutsche Forschungsgemeinschaft (DFG, GZ: SE 1561/1-1) and the kind cooperation of the organ builder Attila Budavári (Pécs Organ-building Manufacture, Hungary).

[1] G. B. Brown, "The vortex motion causing edge tones," Proc. Phys. Soc. London **49**, 493–507 (1937).
[2] A. Powell, "On the edgetone," J. Acoust. Soc. Am. **33**, 395–409 (1961).
[3] W. L. Nyborg, M. D. Burkhard, and H. K. Schilling, "Acoustical characteristics of jet-edge and jet-edge-resonator systems," J. Acoust. Soc. Am. **24**, 293–304 (1952).
[4] O. Coutier-Delgosha, J. F. Devillers, and A. Chaigne, "Edge-tone instability: Effect of the gas nature and investigation of the feedback mechanism," Acta. Acust. Acust. **92**, 236–246 (2006).
[5] A. T. Jones, "Edge tones," J. Acoust. Soc. Am. **14**, 131–139 (1942).
[6] A. Krothapalli and C. Horne, "Recent observations on the structure of an edge-tone flow field," AIAA Paper No. 84-2296 (1984).
[7] A. Krothapalli and W. C. Horne, "Observations on the structure of an edge-tone flowfield," AIAA J. **24**, 1385–1388 (1986).
[8] D. K. Holger, "Fluid mechanics of the edgetone," J. Acoust. Soc. Am. **62**, 1116–1128 (1977).
[9] D. G. Crighton, "The jet edge-tone feedback cycle; linear theory for the operating stages," J. Fluid Mech. **234**, 361–391 (1992).

[10]G. Paál and I. Vaik, "Unsteady phenomena in the edge tone," Int. J. Heat Fluid Flow **28**, 575–586 (2007).
[11]K. Karamcheti, A. B. Bauer, W. L. Shields, G. R. Stegen, and J. P. Woolley, "Some features of an edge-tone flow field," in NASA HQ Conference, (1969) Basic Aeronautical Noise Research NASA Special Publication No. 207, pp. 275–304.
[12]J. W. Coltman, "Jet drive mechanisms in edge tones and organ pipes," J. Acoust. Soc. Am. **60**, 725–733 (1976).
[13]M. S. Howe, *Acoustics of Fluid-Structure Interactions* (Cambridge University Press, Cambridge, 1998).
[14]M. P. Verge, "Aeroacoustics of confined jets—with applications to the physical modelling of recorder-like instruments," Ph.D. thesis, Technical University Eindhoven, Eindhoven, The Netherlands (1995).
[15]M. P. Verge, B. Fabre, W. E. A. Mahu, A. Hirschberg, R. R. van Hasel, A. P. J. Wijnands, J. J. de Vries, and C. J. Hogendoorn, "Jet formation and jet velocity fluctuations in a flue organ pipe," J. Acoust. Soc. Am. **95**, 1119–1132 (1994).
[16]B. Fabre, A. Hirschberg, and A. P. J. Wijnands, "Vortex shedding in steady oscillation of a flue organ pipe," Acust. Acta Acust. **82**, 863–877 (1996).
[17]J. Angster, G. Paál, W. Garen, and A. Miklós, "Effect of voicing steps on the stationary spectrum and attack transient of a flue organ pipe," in Proceedings of the Institute of Acoustics, Institute of Acoustics, St Albans, UK(1997), Vol. **19**, Pt. 5, Book 2 (ISMA '97, Edinburgh), pp. 285–294.
[18]S. Pitsch, "Schneidentonuntersuchungen an einem orgelpfeifenmodell (Edge tone investigations on a foot model of an organ pipe)," MS thesis, Institut für Aero- and Gasdynamik, Fraunhofer Institute for Building Physics, Universität Stuttgart, Stuttgart, Germany (1996).
[19]S. Pitsch, J. Angster, M. Strunz, and A. Miklós, "Spectral properties of the edge tone of a flue organ pipe," in International Symposium on Musical Acoustics, Catgut Acoustical Society, Edinburgh, Institute of Acoustics (1997), Vol. **19**, Pt. 5, Book 2, pp. 339–344.
[20]www.svmtec.de (Last viewed January, 2009).
[21]A. W. Nolle, "Sinuous instability of a planar air jet: Propagation parameters and acoustic excitation," J. Acoust. Soc. Am. **103**, 3690–3705 (1998).
[22]Y. A. Cengel and J. M. Cimbala, *Fluid Mechanics—Fundamentals and Applications (in SI Units)* (McGraw-Hill, New York, 2006).
[23]D. J. Tritton, *Physical Fluid Dynamics* (Van Nostrand Reinhold, New York, 1977).
[24]H. Schlichting, "Laminare strahlausbreitung (Laminar jet propagation)," ZAMM **13**, 260–263 (1933).
[25]W. Bickley, "The plane jet," Philos. Mag. **23**, 727–731 (1939).
[26]S. Yoshikawa, "Harmonic generation mechanism in organ pipes," J. Acoust. Soc. Jpn. (E) **5**, 17–29 (1984).
[27]S. Yoshikawa and K. Arimoto, "Measurement of velocity profiles of the jets issuing from some flue geometries typical of air-jet instruments," in ICA 2001, Rome (2001), Vol. **4**.
[28]R. Janke, "Grundzüge der intonation und strömungsabhilfe (Main features of voicing and lowering disturbances)," http://www.orgel-info.de/ (Last viewed January, 2009).
[29]F. W. White, *Fluid Mechanics* 5th ed. (McGraw-Hill, New York, 2003).
[30]N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments* 2nd ed. (Springer, New York, 1998).
[31]C. Ségoufin, B. Fabre, and M. P. Verge, "Experimental study of the influence of the mouth geometry on sound production in a recorder-like instrument: Windway length and chamfers," Acta Acust. **86**, 649–661 (2000).
[32]A. Bamberger, E. Bänsch, and K. G. Siebert, "Experimental and numerical investigation of edge tones," Z. Angew. Math. Mech. **84**, 632–646 (2004).
[33]C. Ségoufin, B. Fabre, and L. De Lacombe, "Experimental investigation of the flue channel geometry influence on edge-tone oscillations," Acta. Acust. Acust. **90**, 966–975 (2004).
[34]W. L. Nyborg, "Self-maintained oscillation of the jet in a jet-edge system," J. Acoust. Soc. Am. **26**, 174–182 (1954).
[35]Y.-P. Kwon, "Feedback mechanism of low-speed edgetones," J. Acoust. Soc. Am. **104**, 2084–2089 (1998).
[36]J.-C. Lin and D. Rockwell, "Oscillations of a turbulent jet incident upon an edge," J. Fluids Struct. **15**, 791–826 (2001).

Numerical and experimental simulation of the effect of long bone fracture healing stages on ultrasound transmission across an idealized fracture

S. Gheduzzi, S. P. Dodd, and A. W. Miles
Department of Mechanical Engineering, Centre for Orthopaedic Biomechanics, University of Bath, Bath, Avon BA2 7AY, United Kingdom

V. F. Humphrey
Institute of Sound and Vibration Research (ISVR), University of Southampton, Southampton, Hampshire SO17 1BJ, United Kingdom

J. L. Cunningham[a)]
Department of Mechanical Engineering, Centre for Orthopaedic Biomechanics, University of Bath, Bath, Avon BA2 7AY, United Kigndom

(Received 3 March 2008; revised 3 June 2009; accepted 3 June 2009)

The effect of various stages of fracture healing on the amplitude of 200 kHz ultrasonic waves propagating along cortical bone plates and across an idealized fracture has been modeled numerically and experimentally. A simple, water-filled, transverse fracture was used to simulate the inflammatory stage. Next, a symmetric external callus was added to represent the repair stage, while a callus of reducing size was used to simulate the remodeling stage. The variation in the first arrival signal amplitude across the fracture site was calculated and compared with data for an intact plate in order to calculate the fracture transmission loss (FTL) in decibels. The inclusion of the callus reduced the fracture loss. The most significant changes were calculated to occur from the initial inflammatory phase to the formation of a callus (with the FTL reducing from 6.3 to between 5.5 and 3.5 dB, depending on the properties of the callus) and in the remodeling phase where, after a 50% reduction in the size of the callus, the FTL reduced to between 2.0 and 1.3 dB. Qualitatively, the experimental results follow the model predictions. The change in signal amplitude with callus geometry and elastic properties could potentially be used to monitor the healing process.
 [DOI: 10.1121/1.3158938]

PACS number(s): 43.80.Ev, 43.80.Jz, 43.80.Qf, 43.80.Vj [DLM] Pages: 887–894

I. INTRODUCTION

This paper describes a numerical study of the transmission of ultrasonic waves across a simulated transverse bone fracture (as might be observed when using the axial transmission technique) and, in particular, how the progressive healing of the fracture affects the transmission. It indicates that the formation of the callus has a significant effect on the transmitted wave and that the transmission changes in a complex manner through the healing process. Experimental measurements using a simulated fracture in Sawbones® plates are used to confirm the qualitative features of the simulations.

The use of ultrasound for the assessment of musculoskeletal conditions is appealing due to its quantitative potential. Through *in-vitro* experimentation and numerical modeling, advances have been made in the understanding of ultrasound wave propagation along cortical bone. These studies are important for devising techniques for monitoring fracture healing[1–5] and for detecting the onset of osteoporosis.[6–10] In the case of fracture healing monitoring attempts have been made to utilize the change in velocity, measured for a specified transducer separation, to distinguish between ultrasound waves traveling through bone with and without a fracture present. As the velocity is expected to gradually return to the value of the solid bone as the fracture heals this technique could potentially be used to monitor intermediate stages in the repair process[1,11] and, ultimately, predict the end-point of healing.

In the axial transmission technique, the waves propagating along a plate/bone reduce in amplitude as a result of both the absorption in the material and wave energy leaking out of the plate/bone into the surrounding fluid/tissue, a process known as re-radiation. As a consequence of both these effects the re-radiated waves, detected by a receiving ultrasound transducer, reduce in amplitude with distance. Measurements of the amplitude of the re-radiated wave with propagation distance can be used to evaluate the effective attenuation of the ultrasound along the plate. In addition, if there is a discontinuity in the plate/bone, such as a fracture, then the amplitude of the wave transmitted across the discontinuity is reduced. By measuring the re-radiated wave, the fracture transmission loss (FTL) resulting from the discontinuity (i.e., the fracture) can be measured.

a)Author to whom correspondence should be addressed. Electronic mail: j.l.cunningham@bath.ac.uk

Numerical modeling has proved a useful tool for understanding how the velocity and attenuation of ultrasound vary with bone geometry and acoustic frequency,[4,12–14] when measured using an axial transmission technique. In recent two-dimensional (2D) and three-dimensional studies by Protopappas *et al.*,[2,3] the different stages of fracture healing (inflammatory to hard callus) and their effect on ultrasound velocity were simulated, thus providing predictions for the behavior of this particular acoustic parameter during the healing process.

It has recently been demonstrated that ultrasound signal loss has the potential to be a good quantitative indicator of the presence of idealized fractures of varying geometries in bone.[15–17] A number of fracture geometries have been modeled for a particular plate thickness, starting with simple transverse fractures surrounded and filled by water. The water simulates the hematoma present at the inflammatory stage of the fracture healing and the soft tissues surrounding the bone. Numerical models, approximating cortical bone as an isotropic flat plate, compared reasonably well with experimental results[5,12] and simulations provided a good understanding of the relationship between first arrival signal (FAS) velocity and attenuation, plate thickness, and the acoustic frequency. These show that for a small ideal gap the transmitted signal may be reduced by 6 dB or more as a result of the fracture presence.

This work is designed to complement these studies, with experimental and numerical modeling being employed to simulate the axial transmission technique and the amplitude of the ultrasonic waves transmitted across fractures at different stages of the healing sequence. As the fracture heals, the mechanical and geometric properties of the fracture site change;[18] the formation of an external callus and the increasing stiffness of the material in the fracture gap will markedly influence the propagating wave. Hence, the more advanced stages of healing have also been studied using models aimed at replicating a bridging symmetric callus. To simulate the progress of healing, the mechanical and acoustical properties of the callus bridging the fracture have been varied based on reference data gathered from the literature. The changes in ultrasound propagation resulting from the remodeling stage of healing have also been investigated in this study. In order to verify the performance of the numerical model, experimental results were also obtained using Sawbones® plates to simulate the cortical bone and bone cement to simulate the callus.

II. METHODS AND MATERIALS

The modeling studies described here used an axial geometry and parameters based on an experimental facility that has been used previously to study the axial transmission technique.[5,15] This system was also used to obtain the experimental results described in this work.

A. Axial transmission model

The axial transmission technique was modeled using a 2D finite difference code (WAVE2000 PRO, Cyberlogic), the simulation details adopted are tailored to replicate as closely as possible the experimental geometry. In the finite difference numerical models, a rectangular region with dimensions 200×15 mm^2 was used to simulate the water surrounding the cortical bone plate; water was used as an ultrasound mimic for soft tissue and blood in a similar manner as the conditions in the experimental set-up. One ultrasound source (T) and an array of receivers (R) with widths of 20 mm were positioned within the water region at a constant height, $h = 5$ mm, above the plate (see Fig. 1), an initial transducer separation, denoted by x_0, was chosen to avoid signal interference from direct water borne waves. The time and space resolution of these simulations were 0.150 mm and 0.0145×10^{-6} s, respectively.

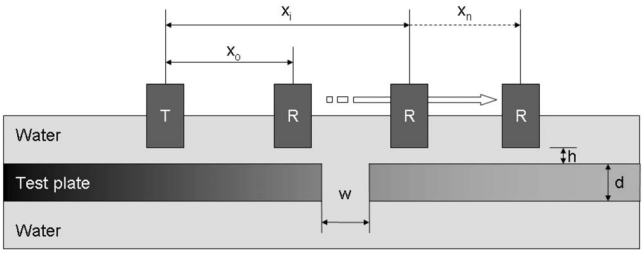


FIG. 1. A geometric representation of the axial transmission apparatus showing the water and test sample, where T and R represent the transmitting and receiving transducers, respectively. x_0, x, d, h, and w are the initial transducer separation, transducer separation, test plate thickness, height of transducer face above plate, and fracture gap width, respectively. The dotted line above point *A* represents the starting position of the measurements.

The source (T) used in the models was configured to output a 4-cycle Gaussian-modulated pulse with 200 kHz central frequency, which simulated the characteristics of the output produced by the experimental transducers. Since the simulated receiving transducers (R) gave an output proportional to the average pressure over the face of the receiver at a particular position and did not have any effect on the wavefield, it was possible to create an overlapping transducer array spaced 1 mm apart to provide data for the FAS amplitude $V(x)$ over a specified propagation distance (x). The FAS amplitude was determined from the peak of the first half cycle of the signal.

B. Fracture geometries for healing study

For the healing process, seven stages were simulated using a 4 mm thick, isotropic, flat bone plate with a 1 mm wide transverse fracture. Figure 2(a) shows the intact bone plate and Figs. 2(b)–2(g) show the geometries used for each of the main stages of the healing processes.[18] Stage 1 [Fig. 2(b)] represents the initial fracture of the plate, which provided information on the inflammatory stage of fracture healing. In this case the gap is filled with water (to simulate blood). Stage 2 [Fig. 2(c)] simulates the initiation of the bridging and ossification process with a symmetric external callus forming to bridge the gap and stabilize the fracture, while cartilage forms at the fracture site. A Gaussian profile with a peak height of 3 mm and a width of 20 mm was used to represent the callus. In this study, height and width refer to the dimensions of the callus and not the medical terms relating to measurement of the whole callus. Subsequently, stages 3 and 4 [Figs. 2(d) and 2(e)] show the continued ossification

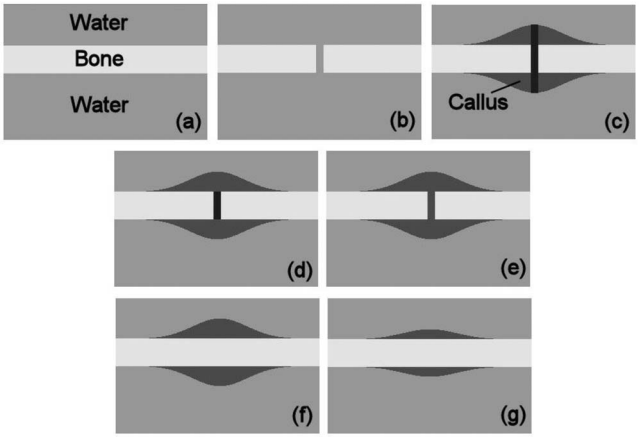


FIG. 2. Modeling pictures of the six stages of secondary healing: (a) an intact bone (light gray). (b) Stage 1: The initial transverse fracture. Stages 2–4: A bridging callus with degrees of ossification, where (c) shows the hard callus material (dark gray) forming the bridge with cartilage filling the gap (black), (d) shows the hard callus material connecting to bridge the gap and cartilage remaining within the fracture gap, and (e) shows the hard callus material throughout the fracture site. Stages 5 and 6: Remodeling, where (f) represents the hard callus material replaced by cortical bone within the fracture gap and (g) the reduction in callus size as the bone returns to original state. Water is used to represent the blood in these simulations.

of the callus and fracture site materials, resulting in a hard callus. In stage 3 the gap is filled with cartilage while in stage 4 it is replaced with stiffer callus material. Finally, stages 5 and 6 [Figs. 2(f) and 2(g)] simulate the remodeling stages of healing where the cortical bone replaces the hard callus material and the external callus reduces in size to return the bone to its original state.

C. Experimental configuration

The *in-vitro* experiments to simulate the healing stages used 6 mm thick Sawbones® plates immersed in a water tank. A baseline measurement was performed over a specified distance on an intact plate; then a 4 mm transverse gap was introduced and the measurement was retaken. While the fracture gap width was maintained, the plate discontinuity was filled using Palacos® R bone cement and the FAS attenuation, over the same distance, was determined. Finally, bone cement was added to the top surface of the plate to simulate a callus. The callus was approximately dome shaped with a height of 4 mm at the center and extended 7.5 mm either side of the center of the fracture site (width of 15 mm). This geometry was maintained across the width of the plate. Figures 3(a)–3(c) provide a representation of the geometrical and mechanical changes.

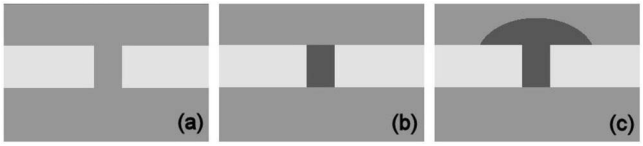


FIG. 3. Modeling pictures representing the three fracture geometries used in the *in-vitro* experiments and 2D simulations. A 6 mm, thick plate, surrounded by water and (a) a 4 mm transverse fracture, (b) bone cement filling/bridging the fracture gap, and (c) a callus is added to the top surface.

In the *in-vitro* experiments, the receiving transducer (R) was moved over the distance (x) at 1 mm intervals using a linear stepper motor rig with the separation being measured by an optical encoder (accuracy of $\pm 1\%$). The received signals were filtered, amplified, and captured using a digital oscilloscope and averaged over 128 waveforms. As in the simulations, the amplitude of the first peak at each separation interval was used to calculate the FAS attenuation, with the experimental data recorded for three separate runs and then averaged.

These experimental measurements were also simulated numerically using an appropriate callus geometry.

D. Simulated material, elastic, and acoustic properties

Material and acoustic properties used in the simulations were estimated from data in the literature and are shown in Table I. In the healing stage models, human cortical bone was represented by the properties, $\rho = 1850$ kg/m^3, $V_L = 4000$ m/s, and $V_S = 1800$ m/s ($E = 16.5$ GPa), which were used in previous studies.[5,13] Cartilage properties ($\rho = 1050$ kg/m^3, $V_L = 1775$ m/s, and $V_S = 946$ m/s, giving $E = 2.45$ GPa) were estimated using a velocity range for hu-

TABLE I. Material, elastic, and acoustic properties used in numerical simulations. Data taken from the literature are referenced. The remaining values were estimated or calculated from referenced data using standard equations.

Material	ρ (kg m^{-3})	E (GPa)	σ	V_L (m s^{-1})	V_S (m s^{-1})
Cortical bone	1850[a]	16.45	0.37	4000[a]	1800[a]
Water/blood	1000[a]	⋯	⋯	1500[a]	⋯
Cartilage	1050[b]	2.45	0.3[c]	1775[b]	946[b]
Callus material 1	1600	5	0.3[c]	2050	1095
Callus material 2	1600	10	0.3[c]	2900	1551
Callus material 3	1600	15	0.3[c]	3553	1897
Sawbones®	1700[d]	11.3	0.35	3300[d]	1600[d]
Palacos® R bone cement	1281[e]	3.25	0.33	2750	1380

[a]Reference 5.
[b]Reference 19.
[c]Reference 20.
[d]Reference 15.
[e]Reference 23.

man specimens,[19] and Poisson's ratio given by Claes and Heigele.[20] A bovine cartilage density of 1050 kg/m^3 reported by Joseph *et al.*[21] was assumed to be a good estimate for the human equivalent, since data on human cartilage are scarce. Human and bovine cartilages have similar constituents (60%–80% water by weight[22]), therefore resulting in similar densities. Material damping values for these materials have not been included due to the lack of reliable data at 200 kHz.

The density and acoustic properties of the callus were difficult to predict due to the ossification process.[20] Therefore, a range of Young's modulus values (5–15 GPa) were modeled for this callus material (Table I), with Poisson's ratio based on data by Claes and Heigele.[20] These values lie in between the simulated mechanical properties of cartilage and cortical bone. In a similar manner, the density of the hard callus material will lie between cartilage and cortical bone; however, for this study it was estimated to be closer to cortical bone (1600 kg/m^3).

For the experimental section of the study, Sawbones® plates and Palacos® R bone cement were used. The modeling of these experimental results used appropriate data based on values from the literature. Sawbones® properties were taken from a similar study,[5] while the density of the bone cement has been measured previously by Armstrong *et al.*[23] Due to the amount of acrylic present in this particular bone cement[24] and the measured density, the acoustic velocities were estimated to be similar to Perspex. Therefore, these values were based on the material database in the WAVE2000 PRO software package.

E. Fracture loss calculation

The amplitude data (for the FAS) was plotted as a sound pressure level (SPL), in decibels, versus transducer separation (m) using

$$\mathrm{SPL}(x) = 20\log(V(x)/V(x_0)), \tag{1}$$

where $V(x_0)$ is the signal amplitude at $x=x_0$ and $V(x)$ is the signal amplitude at each measurement position x.

The difference in SPLs at a given measurement position (x) was used previously[15,16] to represent the change in FAS amplitude caused by a discontinuity and was expressed as a FTL, calculated using

$$\mathrm{FTL}(x) = -(\mathrm{SPL}_F(x) - \mathrm{SPL}_I(x)), \tag{2}$$

where the subscript I denotes an intact specimen and F denotes a fractured specimen. These parameters are used throughout the paper. With the notation of Eq. (2), a reduction in signal as a result of the fracture will give rise to a positive FTL.

III. RESULTS

A. Effect of different healing stages on the FAS amplitude

Typical experimentally measured and numerically simulated signals are presented in Figs. 4(a) and 4(b), respectively. Figure 4(b) shows the complex nature of the received signals with more than one wave contributions resulting in interference effects between the different contributions. The nature of the waves received when using the axial transmission technique has been discussed previously.[13,5,10] This has shown that the first arrival travels at the velocity of a compressional wave or S_0 mode depending on the range and plate thickness. In this work the amplitude of the first peak of the transmitted waveform [Fig. 4(a)] is used for simplicity.

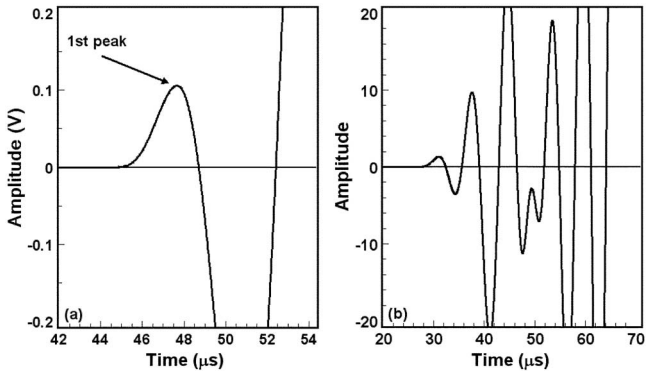
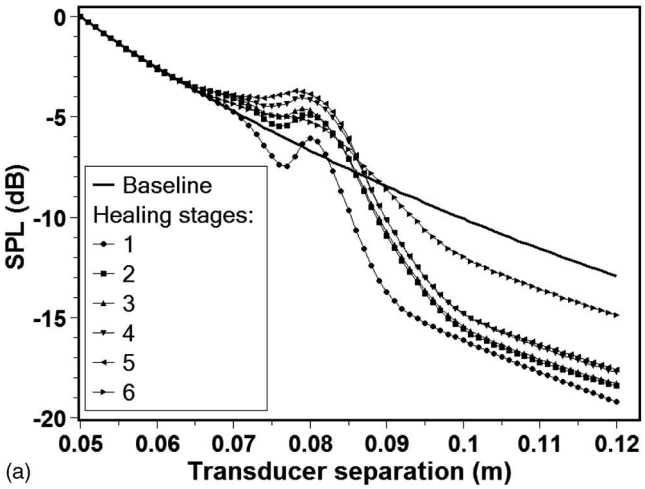
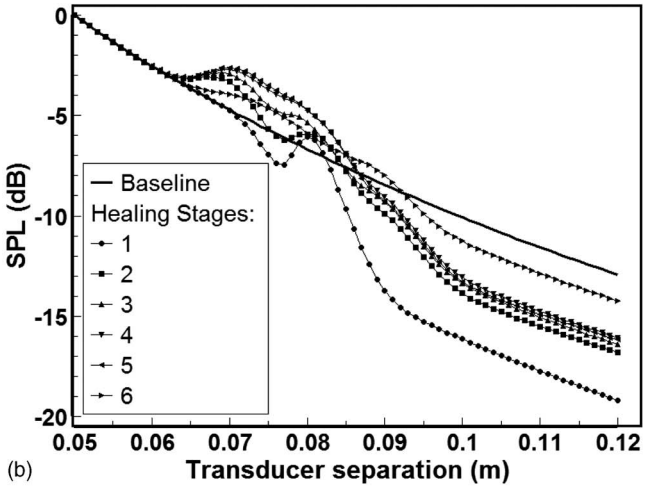


FIG. 4. Typical experimental (a) and simulated (b) signals.

Figure 5(a) shows the effect of the simulated healing stages 1–6 on the FAS amplitude of an ultrasound wave propagating along a 4 mm bone plate with a 1 mm transverse fracture gap. With reference to Figs. 5(a)–5(c), the first fracture interface is positioned at a transducer separation of 0.07 m. In this particular study, the net signal loss, or FTL, was considered. This parameter gave the difference in SPL between the intact bone plate and the subsequent healing stages of the fractured bone plate at a particular transducer separation. Figure 5(a) shows the results obtained using callus material 1 (E=5 GPa); the geometry and mechanical properties of each healing stage resulted in a different shape to the corresponding SPL curves. At the maximum transducer separation (120 mm), the difference in signal amplitude between the subsequent healing stages and the baseline value varied significantly. These differences expressed as both a percentage change from the baseline data and as a FTL are presented in Table II.

The simulation of stage 1, which represented the inflammatory stage of healing, produced a significant drop in the signal amplitude compared to the intact plate [Fig. 5(a)], resulting in a large FTL. When the bridging callus was introduced in stage 2, the FTL value was reduced relative to stage 1, i.e., the signal amplitude increased. Stages 2 and 3 represented a bridging callus and little change in FTL was seen at stage 3 as the external callus formed to bridge the gap. A subsequent change to a hard callus (stage 4), i.e., when the ossification had occurred throughout the fracture site, resulted in a small decrease in the loss of signal amplitude, and hence FTL, relative to stage 3. However, the beginning of the remodeling stage (stage 5) caused no significant change in the FTL value compared to stage 4. Only when the callus reduced in size in stage 6 did the FTL change significantly.





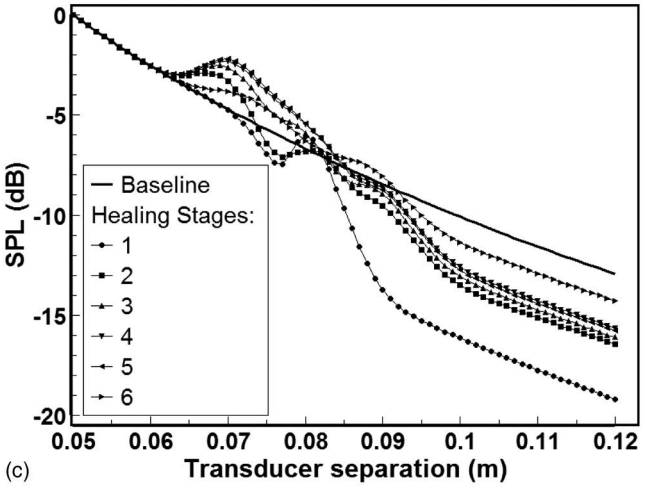


FIG. 5. Numerical modeling results of the SPL versus transducer separation for the six healing stages when Young's modulus value corresponds to callus materials (a) 1 (5 GPa), (b) 2 (10 GPa), and (c) 3 (15 GPa). FTL calculations where preformed using data taken at 120 mm. The first fracture interface is positioned at a transducer separation of 0.07 m.

B. Effect of increasing Young's modulus of the callus material

The effect of increasing Young's modulus (E) of the callus material on the FTL values produced by the simulations are demonstrated in Figs. 5(a)–5(c). Table II summarizes the percentage change in amplitude compared with the baseline data and shows the FTL calculated from these curves for each Young's modulus value used (callus materials 1–3, 5–15 GPa). A callus with a Young's modulus of 10 GPa produced a larger drop in FTL between stages 1 and 2 [Fig. 5(b)], compared to the same fracture geometry with half the stiffness, $E=5$ GPa [Fig. 5(a)]. Although there exists a continuous fall in FTL from stages 2 to 5, the change is relatively small, in a similar way to that obtained for $E=5$ GPa. A much larger drop in FTL occurred from stages 5 to 6 than existed for original Young's modulus. When Young's modulus of the callus material was increased further to 15 GPa [Fig. 5(c)], the same pattern was produced with similar FTL values to those obtained for 10 GPa.

TABLE II. A comparison between the FTL calculated from numerical modeling data of the signal loss produced by healing stages 1–6 and increasing Young's modulus of the callus material. The change in signal amplitude (compared with the baseline data) is also expressed as a percentage and given in parentheses.

	Callus material		
Healing stage	1 (5 GPa)	2 (10 GPa) FTL (dB)	3 (15 GPa)
1	6.3 (−52%)	6.3 (−52%)	6.3 (−52%)
2	5.5 (−47%)	3.9 (−36%)	3.5 (−33%)
3	5.4 (−46%)	3.5 (−33%)	3.2 (−31%)
4	4.8 (−42%)	3.2 (−31%)	2.7 (−27%)
5	4.7 (−42%)	3.3 (−32%)	2.9 (−28%)
6	2.0 (−21%)	1.3 (−14%)	1.3 (−14%)

C. *In-vitro* experiments of changing the fracture geometry and mechanical properties of a fracture site

Corresponding experimental FAS amplitude results of changing the shape and geometry of the fracture site for a Sawbones® bone plate mimic with a 4 mm transverse fracture are shown in Fig. 6(a). For the results presented in Fig.

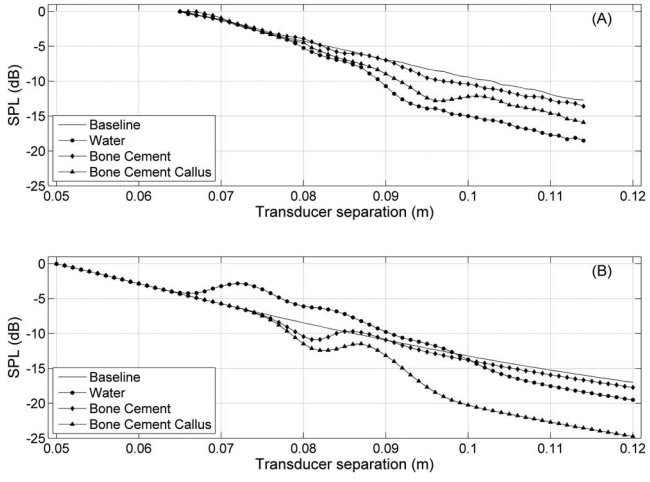


FIG. 6. Results of the SPL versus transducer separation measurements for (a) *in-vitro* experiments investigating the effect of fracture geometry and mechanical properties on signal amplitude, and (b) simulations of the *in-vitro* experiments. A 6 mm thick Sawbones® plate with a 4 mm transverse fracture was used for this work. The first fracture interface is positioned at a transducer separation of 0.07 m.

TABLE III. A comparison between the FTL measured from *in-vitro* experiments and predicted by numerical modeling of the signal loss produced by three different fracture geometries simulated on a Sawbones® plate. The change in signal amplitude (compared with the baseline data) is also expressed as a percentage and given in parentheses.

	FTL (dB)	
Gap material	Measured	Predicted
Water	5.9 (−49%)	7.5 (−58%)
Bone cement	0.2 (−3%)	0.7 (−8%)
Bone cement callus	3.0 (−29%)	2.3 (−23%)

6 the first fracture interface was positioned at a transducer separation of 0.07 m. A large drop in signal amplitude is seen for the inflammatory stage, equivalent to stage 1 in the numerical modeling. When an intermediate material, Palacos® R bone cement, was substituted for the water in the gap, the signal loss was reduced significantly, i.e., the signal amplitude at large separation was only slightly lower than the baseline data, giving a small FTL. Adding a callus to the top surface of the bone mimic plate resulted in increased signal loss and hence an increase in FTL. The experimental uncertainty associated with the experimental baseline and callus measurements of amplitude was approximately 10%.

D. Modeling of the callus experiment

Figure 6(b) shows the results of modeling the *in-vitro* experiment. The model predicted similar changes in the signal amplitude at large separations for the different fracture geometries used in the experiments. When the material in the fracture gap was given the properties of the cement, the signal loss was small compared to the baseline data, but the addition of a callus caused the signal loss to increase and resulted in a larger FTL (Table III). One significant difference between the experimental data and the numerical modeling was the shape of characteristic curve for the callus geometry. In the model, a large peak in the reradiated amplitude was produced; this was not present in the experimental data.

IV. DISCUSSION

Previous studies of the relative change in velocity as a transverse fracture healed[1,2,11] have shown that, after the initial fracture, the velocity tended to increase steadily toward the baseline (intact) value. The models of a healing transverse fracture used in this study suggest that the relative change in FAS amplitude displays a different variation with the healing stages when compared with the velocity changes. It is also interesting to compare the magnitude of the changes observed in average velocity and transmission amplitude. For example, for a 2 mm fracture gap and 20 mm transducer spacing, the maximum reduction in average velocity reported by Protopappas *et al.*[2] was of order 20%. For the attenuation measurements reported here for a gap of 2 mm the reductions in transmission are of order 50%. Velocities will in general be easier to measure accurately, but as the transducer separation increases the change in average velocity reduces unlike the FTL.

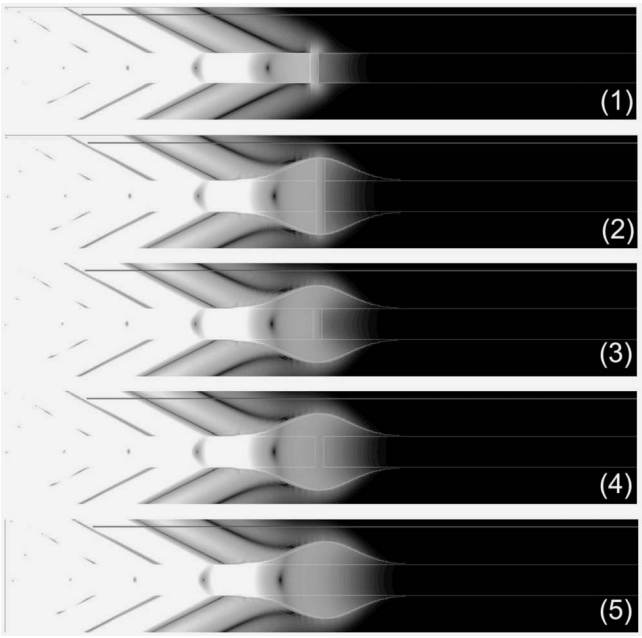


FIG. 7. Numerical simulation snapshots of acoustic pressure for healing stages 1–5. At stage 1 the fracture gap is filled with water, stage 2 represents the initiation of bridging and ossification with a symmetrical noncontinuous callus and cartilage at the fracture site, and stages 3 and 4 represent the continued ossification of callus, the difference being the material filling the fracture gap which is modeled as cartilage in stage 3 and callus in stage 4. The front of the acoustic pulse has just passed the fracture.

For a relatively low modulus callus ($G=5$ GPa), the changes in the FAS amplitude were not steady with the healing process stages. The change in geometry, i.e., introduction of a bridging callus (stages 1 and 2), and the low stiffness of the callus material did not result in a significant decrease in the FTL. Similarly, the change in mechanical properties, as the fracture geometry remained constant (stage 2–5), did not result in significant decreases in the signal loss. This would suggest that the mechanical changes within the callus do not greatly affect the measurements. The most apparent change occurred in the simulated remodeling stage (stage 6), where the loss was approximately halved (see Table II).

This picture changed when the stiffness of the callus material was increased to 10 GPa. Essentially, small increases toward the baseline result (decrease in FTL) were observed for data taken in stages 2–5, while a marked change was observed between stages 1 and 2 and between stages 5 and 6. The small variation in amplitude for stages 2–5 reinforced the idea that the mechanical changes within the gap do not greatly affect the measurements. A further increase in callus stiffness (to 15 GPa) did not significantly alter this behavior.

The reduction in fracture loss as the callus is added (from stages 1 to 2) indicates that the callus helps to guide the acoustic radiation from one side of the fracture to the other, effectively acting as a waveguide. This effect is shown in Fig. 7 which shows snapshots of the numerical simulations for healing stages 1–5 at a time when the front of the pulse has just passed the fracture. It should be noted that from stage 2 the pressure wave spreads out in to callus, with similar pressures in the callus and bone. This effect increases

as the modulus of the callus increases due to the better matching of the bone to the callus. At the same time, however, the callus acts as an additional load on the bone giving rise to an increased re-radiation from the fracture site. This effect then appears to dominate the loss, so changes in the fracture gap material modulus have relatively little effect on the fracture loss (stages 2–5). As the density of the callus remains constant, the mass loading will not vary with callus material. As a result, the loss only starts to approach the baseline as the callus reduces in size (stages 6 and 7).

Some experiments were performed to simulate the effect of changing the mechanical properties and geometry of a fracture site, in this case using flat plates. These experiments provided some useful insights, even though they do not use the same geometries as the numerical modeling. Increasing the stiffness of the simulated healing tissue, i.e., substituting bone cement for water in the fracture gap, resulted in a significant decrease in signal loss (higher transmission). The gap filler reduces the mismatch at the two gap interfaces and increases the transmission across the fracture zone. This pattern is also seen in the modeling of these experiments. However, adding a bone cement callus to the top surface across the fracture resulted in an increased signal loss compared to just the gap filler alone. This confirms the concept that the callus itself can give rise to strong scattering/re-radiation and a reduction in signal transmission across the fracture and this may have some clinical relevance depending on the extent of callus formation during healing.

In terms of sensitivity of the signal amplitude measurement to changing mechanical and geometrical properties at the fracture site, these models and experiments suggest that only three changes are detectable. The first change from stages 1 to 2 (initial inflammatory condition to the formation of a callus) is a result of both the geometrical and mechanical properties of the fracture site. However, a stiffer callus material creates a larger relative signal amplitude change at this stage. Mechanical changes within the callus that occur from stages 2 to 5 though potentially clinically significant cannot be easily detected with this experimental set-up. It is possible that the change in signal amplitude from stages 5 to 6 may be proportional to the callus dimensions. Therefore, the next detectable change would occur in the remodeling stage after a significant reduction in the callus size has taken place. The final change occurs when the original geometry of the bone is regained.

In these models and experiments, only the transverse case was considered. A number of physical properties need to be investigated to provide a more complete picture of the wave interaction at different healing stages. The thickness of the bone plate at a particular acoustic frequency has been shown to influence FAS velocity and attenuation measurements; therefore, modeling of different geometries, such as oblique fractures, and different plate thicknesses may also provide important insights on the modalities of ultrasound propagation in bone. Other properties, such as cortical bone porosity and material damping, may also have an effect on the relative amount of ultrasound signal loss detected.

V. CONCLUSIONS

The results presented confirm that the addition of a discontinuity to a cortical bone model, such as a transverse fracture, results in a reduction in the reradiated FAS amplitude beyond the discontinuity. The use of a numerical finite difference model has enabled the effects of the different stages of the fracture healing process on the acoustic transmission to be studied. These show that the changes can be complex, and do not vary steadily with the healing stage.

Specifically the largest changes in the transmitted amplitude are observed to occur from stages 1 to 2, the initial inflammatory condition to the formation of a callus, and from stages 5 to 6 in the remodeling stage after a significant reduction in the callus has taken place. The relative significance of these two stages appears to be dependant on Young's modulus selected for the callus, with the former dominating for higher moduli.

The progressive changes in the modulus of the material filling the fracture appears to produce less significant changes in the FTL, irrespective of the callus modulus. These results imply that it is important to consider the effect of the callus when considering the changes that occur in acoustic transmission across a fracture through the fracture healing process. The callus itself appears to be a strong scatterer/source of re-radiation, presumably due to its mass loading of the bone plate.

These numerical studies have been confirmed by *in-vitro* experiments performed using flat plates. These demonstrated that substituting bone cement for water in the fracture gap resulted in a large decrease in the FTL, i.e., better signal transmission, compared with the water gap. However, adding a bone cement callus to the top surface across the fracture resulted in an increased signal loss compared to just the gap filler alone.

The results indicate that significant changes do occur, both in the re-radiation from the fracture/callus and transmission loss across the fracture site, through the fracture healing sequence. This change in the signal amplitude from an inflammatory healing stage to callus formation stage or to the callus remodeling stage could potentially be used to quantify the healing process. In practice the amplitude values without a fracture would be obtained from measurements of the non-fractured contraleteral bone. The aim of the clinical measurement would then be to evaluate the evolution of the FTL from the fractured bone with time as healing progressed and as it approached, but did not necessarily coincide with, the FTL value for the non-fractured contralateral. For example, a fracture could be deemed sufficiently healed for unsupported weight-bearing if the FTL of the fractured bone was within a certain percentage of that of the contralateral.

ACKNOWLEDGMENT

The authors are grateful to Action Medical Research for funding this work.

[1]V. C. Protopappas, D. A. Baga, D. I. Fotiadis, A. C. Likas, A. Papachristos, and K. N. Malizos, "An ultrasound wearable system for the monitoring and acceleration of fracture healing in long bones," IEEE Trans. Biomed. Eng. **52**, 1597–1608 (2005).

[2]V. C. Protopappas, D. I. Fotiadis, and K. N. Malizos, "Guided ultrasound wave propagation in intact and healing long bones," Ultrasound Med. Biol. **32**, 693–708 (2006).
[3]V. C. Protopappas, I. C. Kourtis, L. C. Kourtis, K. N. Malizos, C. V. Massalas, and D. I. Fotiadis, "Three-dimensional finite element modelling of guided ultrasound wave propagation in intact and healing long bones," J. Acoust. Soc. Am. **121**, 3907–3921 (2007).
[4]G. Barbieri, C. H. Barbieri, P. S. De Matos, C. A. Pelá, and N. Mazzer, "Ultrasonometric evaluation of bone healing: Experimental study using a model of diaphyseal transverse osteotomy of sheep tibiae," Ultrasound Med. Biol. **32**, 875–882 (2006).
[5]S. P. Dodd, J. L. Cunningham, A. W. Miles, S. Gheduzzi, and V. F. Humphrey, "Ultrasonic propagation in cortical bone mimics," Phys. Med. Biol. **51**, 4635–4647 (2006).
[6]T. Chen, P.-J. Chen, C.-S. Fung, C.-J. Lin, and W.-J. Yao, "Quantitative assessment of osteoporosis from the tibia shaft by ultrasound techniques," Med. Eng. Phys. **26**, 141–145 (2004).
[7]G. Haïat, F. Padilla, R. Barkmann, S. Kolta, C. Latremouille, C.-C. Glüer, and P. Laugier, "In vitro speed of sound measurement at intact human femur specimens," Ultrasound Med. Biol. **31**, 987–996 (2005).
[8]A. Tatarinov, N. Sarvazyan, and A. Sarvazyan, "Use of multiple acoustic wave modes for assessment of long bones: Model study," Ultrasonics **43**, 672–680 (2005).
[9]D.-a. Ta, K. Huang, W.-q. Wang, Y.-Y. Wang, and H. L. Le, "Identification and analysis of multimode guided waves in tibia cortical bone," Ultrasonics **44**, E279–E284 (2006).
[10]P. Moilanen, P. H. F. Nicholson, V. Kilappa, S. Cheng, and J. Timonen, "Measuring guided waves in long bones: Modelling and experiments in free and immersed plates," Ultrasound Med. Biol. **32**, 709–719 (2006).
[11]E. Maylia and L. D. M. Nokes, "The use of ultrasound in orthopaedics—A review," Technol. Health Care **7**, 1–28 (1999).
[12]P. H. F. Nicholson, P. Moilanen, T. Karkkainen, J. Timonen, and S. Cheng, "Guided ultrasonic waves in long bones: Modelling, experiment and *in vivo* application," Physiol. Meas **23**, 755–768 (2002).
[13]E. Bossy, M. Talmant, and P. Laugier, "Effect of bone cortical thickness on velocity measurements using ultrasonic axial transmission: A 2D simulation study," J. Acoust. Soc. Am. **112**, 297–307 (2002).
[14]E. Bossy, M. Talmant, and P. Laugier, "Three-dimensional simulations of ultrasonic axial transmission velocity measurements on cortical bone models," J. Acoust. Soc. Am. **115**, 2314–2324 (2004).
[15]S. P. Dodd, J. L. Cunningham, A. W. Miles, S. Gheduzzi, and V. F. Humphrey, "An in-vitro study of ultrasound signal loss across simple fractures in cortical bone mimics and bovine cortical bone samples," Bone **40**, 656–661 (2007).
[16]S. P. Dodd, A. W. Miles, S. Gheduzzi, V. F. Humphrey, and J. L. Cunningham, "Modelling the effects of different fracture geometries and healing stages on ultrasound signal loss a long bone fracture," Comput. Methods Biomech. Biomed. Eng. **10**, 371–375 (2007).
[17]S. P. Dodd, J. L. Cunningham, A. W. Miles, S. Gheduzzi, and V. F. Humphrey, "Ultrasound transmission loss across transverse and oblique bone fractures: An *in-vitro* study," Ultrasound Med. Biol. **34**, 454–462 (2008).
[18]R. B. Martin, D. B. Burr, and N. A. Sharkey, *Skeletal Tissue Mechanics* (Springer-Verlag, New York, 1998), p. 67.
[19]S. G. Patil, Y. P. Zheng, J. Y. Wu, and J. Shi, "Measurements of depth-dependence and anisotropy of ultrasound speed of bovine articular cartilage *in vitro*," Ultrasound Med. Biol. **30**, 953–963 (2004).
[20]L. E. Claes and C. A. Heigele, "Magnitudes of local stress and strain along bony surfaces predict the course and type of fracture healing," J. Biomech. **32**, 255–266 (1999).
[21]D. Joseph, W. Y. Gu, X. G. Mao, W. M. Lai, and V. C. Mow, "True density of normal and enzymatically treated bovine articular cartilage," Trans. Annu. Meet. - Orthop. Res. Soc. **24**, 642 (1999).
[22]S. Saarakkala, M. S. Laasanen, J. S. Jurvelin, K. Törrönen, M. J. Lammi, R. Lappalainen, and J. Töyräs, "Ultrasound indentation of normal and spontaneously degenerated bovine articular cartilage," Osteoarthritis Cartilage **11**, 697–705 (2003).
[23]M. S. Armstrong, R. F. Spencer, J. L. Cunningham, S. Gheduzzi, A. W. Miles, and I. D. Learmonth, "Mechanical characteristics of antibiotic-laden bone cement," Acta Orthop. Scand. **73**, 688–690 (2002).
[24]J. Algers, F. H. J. Maurer, M. Eldrup, and J.-S. Wang, "Free volume and mechanical properties of Palacos® R bone cement," J. Mater. Sci.: Mater. Med. **14**, 955–960 (2003).

Blind location and separation of callers in a natural chorus using a microphone array

Douglas L. Jones
Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 1308 W. Main Street, Urbana, Illinois 61801

Rama Ratnam[a)]
Department of Biology, University of Texas at San Antonio, One UTSA Circle, San Antonio, Texas 78249

(Received 17 December 2008; revised 30 April 2009; accepted 27 May 2009)

Male frogs and toads call in dense choruses to attract females. Determining the vocal interactions and spatial distribution of the callers is important for understanding acoustic communication in such assemblies. It has so far proved difficult to simultaneously locate and recover the vocalizations of individual callers. Here a microphone-array technique is developed for blindly locating callers using arrival-time delays at the microphones, estimating their steering-vectors, and recovering the calls with a frequency-domain adaptive beamformer. The technique exploits the time-frequency sparseness of the signal space to recover sources even when there are more sources than sensors. The method is tested with data collected from a natural chorus of Gulf Coast toads (*Bufo valliceps*) and Northern cricket frogs (*Acris crepitans*). A spatial map of locations accurate to within a few centimeters is constructed, and the individual call waveforms are recovered for nine individual animals within a 9×9 m^2. These methods work well in low reverberation when there are no reflectors other than the ground. They will require modifications to incorporate multi-path propagation, particularly for the estimation of time-delays.
 [DOI: 10.1121/1.3158924]

PACS number(s): 43.80.Ka, 43.60.Fg, 43.60.Jn [MO] Pages: 895–910

I. INTRODUCTION

The determination of individual interactions in animal choruses is of wide interest to behavioral researchers. Participants in a chorus can dynamically adjust spacing and acoustic call timing to influence behaviors such as mate choice, defense of territory and group cohesion (Kroodsma *et al.*, 1983; Sullivan *et al.*, 1995; Gerhardt and Huber, 2002; Simmons *et al.*, 2002). But the number of participants, their spatial distribution, and the overlap between calls makes it difficult to precisely locate the individuals and separate their voices. Consequently, spatial and temporal characteristics of a chorus are typically studied in smaller subsets of the chorus, where it may be easier to visually locate callers and record the sound of one or a few individuals. What would be of benefit to researchers is a technique for mapping a chorus that simultaneously locates callers and separates their voices over a much larger scale. Such a spatio-temporal map would make it possible to address questions on large-scale chorus dynamics. The research reported here is a step in the direction of chorus mapping. Building on past work and new research, it outlines techniques for locating callers and separating their voice using a microphone array. The application discussed here is a frog chorus, but the techniques can be applied to other chorusing species as well.

A. Anuran choruses

Vocally communicating anurans such as frogs and toads congregate in dense choruses around bodies of water and vocalize to attract females. This is a lek-like breeding system (Bradbury, 1981) where males contribute sperm but do not otherwise control a resource. Females are therefore free to choose mates. Anuran choruses are usually heterospecific and their notable feature is high call density. Numerous individuals from many different species may be calling at the same time with a high degree of temporal and spectral overlap. While species calls are stereotypical and sufficiently distinct in their temporal and spectral features to reproductively isolate the species (Blair, 1958; Bogert, 1960), call overlap and the presence of large numbers of callers in close proximity can give rise to acoustic jamming and masking interference (Ehret and Gerhardt, 1980; Narins, 1982; Gerhardt and Klump, 1988; Schwartz and Wells, 1983a, 1993b; Schwartz and Gerhardt, 1995; Wollerman, 1999).

At the individual level, communication in dense choruses presents several challenges to signalers (males) and receivers (females). Females approaching the chorus must detect, locate, and identify calling conspecific males in the presence of masking interference. Thus, a target call must be sufficiently intense and well-separated from interfering sources. On the other hand, males are confronted with the problem of defending their acoustic space so that they may time their calls to be heard above the background (see Gerhardt and Huber, 2002, for a review). Thus, location and spatial separation, in combination with call intensity, timing,

[a)]Author to whom correspondence should be addressed. Electronic mail: rama.ratnam@utsa.edu

and overlap, are important parameters that may be actively controlled to maximize mate attraction and selection.

Earlier studies have shown that frogs have evolved strategies for utilizing time, frequency, and space in energetically efficient ways to minimize wasteful calling. Most of this evidence comes from small-scale studies involving a few frogs (Zelick and Narins, 1985; Gerhardt *et al.*, 1989; Wilczynski and Brenowitz, 1988; Brush and Narins, 1989; Grafe, 1997; Simmons *et al.*, 2006, 2008; Schwartz, 2001) but there are little data on the spatial and temporal structures of natural choruses involving large numbers of callers. Evidence from these studies point to interactions between local groups of callers with males typically paying greater attention to the calls of nearest neighbors (Brush and Narins, 1989; Schwartz, 2001). Such interactions will ncessarily weaken with increasing inter-male separation because of sound attenuation. Wagner and Sullivan (1992) suggested that males move when the number density of callers increases while preferring to remain stationary when densities are low. But little is known about the adjustment in individual call timing as a function of spacing (see Brush and Narins, 1989, for an exception). Such data are not easy to obtain without physically moving calling individuals and disrupting the natural scene. These concerns demonstrate the need for wider noninvasive spatial sampling in conjunction with call extraction.

At the chorus level a different set of questions arises particularly with regard to sexual selection. Broadly, anurans can be classified as prolonged or explosive breeders (Wells, 1977). Prolonged breeders have a long breeding season and the number density of males is relatively low. Explosive breeders, on the other hand, have a short breeding season and sustain higher densities. Both types of breeders can be sympatric as is the case for the two species considered here. A motivation for measuring chorus density, and to a lesser extent chorus size, has been the determination of operational sex ratios and female behavior, and male mating success. A major theoretical direction has been to determine how these factors influence sexual selection.

The data on chorus density are highly variable, being governed by physical environmental parameters and by species characteristics. Consequently density of callers and chorus size are subject to large variability between days in a given season, between seasons, between locations, and between species (see Wagner and Sullivan, 1992). As stated before, gross chorus parameters have usually been measured from the point of view of studying sexual selection and not with the aim of taking a census (for instance, see Gerhardt, 1994; Friedl and Klump, 2005; Murphy, 1994, 2003; Stewart and Pough, 1983). Most of the available data are restricted to a single species and to breeding sites that can be rapidly covered for censusing, typically in one night. By necessity this restricts the size of the choruses that can be covered. In particular, census data for heterospecific choruses are sparse.

For the two species that are studied here, Blanchard's cricket frog (*Acris crepitans blanchardi*, a prolonged breeder) and the Gulf Coast toad (*Bufo valliceps*, an explosive breeder), some data are available. For cricket frogs, Perrill and Shepherd (1989) reported that 10–30 marked males were observed in small ponds over a 3 year period. Their separation was almost always greater than 50 cm, they called for durations ranging from 3 to 6 h, and they typically remained at the same position, sometimes returning to the same location on different nights. Unfortunately no data on the size of the ponds are available, nor is it known whether all marked individuals called in a chorus bout, or indeed whether unmarked males were present.

Wagner and Sullivan (1992), and Sullivan and Wagner (1988) observed Gulf Coast toads over a period of four seasons in two locations. They reported a mean number of toads ranging from 4 to 25 per nightly chorus, with densities ranging from 0.01 to 0.5 toads/m of shoreline. Nearest-neighbor distances were about 5.5 m with large variability (standard deviation of 5.1 m). The number of chorus participants ranged from 2 to 65 males/night and was highly variable between days in a given season, between seasons, and between locations. Males generally remained stationary, moving only when the density increased.

While direct observation and a manual census are often necessary and are unlikely to be replaced, they are arduous and time-consuming, and limited in spatial coverage. Thus, there is a pressing need for locating chorusing frogs if only to assess densities and numbers on a much larger scale, and over multiple species. Likewise, in the temporal domain, it is difficult to manually determine how many individuals are calling at any given time and to determine the call densities (number of callers per unit time). Such data are not readily available and form a major motivation for this work. An automated procedure that does not require direct observation and manual counting would be a valuable adjunct to ongoing research. Thus, the case for automatically localizing and separating callers stems from these observations. The major question is: can it be accomplished? The problem is non-trivial especially when the chorus density is high and there are multiple overlapping heterospecific callers. To further understand the issues, we review some of the methods that can be employed.

B. Microphone-array techniques

A chorus can be mapped acoustically using a spatially dispersed microphone array by following a two-step process. First, the spatial position of each individual caller (source) is determined from the differences in the time of arrival of sounds at the microphones (a procedure akin to triangulation). Second, the known location of each source is used to estimate a steering vector to that source. Then an adaptive spatial filter steers to the selected source and recovers it while suppressing all other sources.

1. Source localization

Many algorithms have been proposed for source localization using intermicrophone time delays (see Carter, 1981). For airborne signals assuming a constant velocity of sound, the time-of-arrival differences (TOADs) are proportional to the differences in source-to-microphone range (range differences), with the locus of points that satisfy a constant range difference being described by hyperboloids (van Etten, 1970). In this method the sensors form the foci of the hyper-

boloids and the source lies at their intersection. In general, a minimum of three or four sensors are required for locating a source in two dimensions, and four or five sensors are required for locating in three dimensions[1] with redundant measurements being used to provide improved accuracy when measurements are noisy (Hahn and Tretter, 1973). The method of hyperbolic range-difference location originated in the Loran navigation system and suffers from several drawbacks. It is a two-step process that requires an estimate of the TOADs before solving for the locations. Thus, it is suboptimal. For instance, when multi-path propagation is present, the arrival-time estimates are not accurate (but see Hahn and Tretter, 1973, for a maximum-likelihood etimator). Further, it does not have a closed-form solution, it is computationally expensive, and it is not amenable to statistical analysis when measurement noise is present.

Several variations of the method have been proposed to overcome some of these drawbacks including a simpler localization scheme where the source location forms the focus of a general conic (Schmidt, 1972). To incorporate measurement noise a generalized least-squares location estimator based on the approach of Schmidt (1972) was proposed by Delosme *et al.* (1980). And to overcome the computational difficulties engendered by manipulating hyperboloids, closed form localization based on spherical methods have been proposed by Schau and Robinson (1987), and Smith and Abel (1987). Additionally, several frequency-domain methods have been proposed. For the general case of multiple sources and sensors, and when the spectral density matrices of the sources and noise are known, Wax and Kailath (1983) extended the single-source results due to Hahn and Tretter (1973) by deriving a maximum-likelihood localizer and the Cramer–Rao lower bound on the error covariance matrix. The method includes the case of multi-path propagation and is a bank of beamformers each directed toward a particular source. More recently Mohan *et al.* (2008) derived a localizer for multiple sources using small arrays (where the number of sensors may be less than the number of sources). They exploit the sparse time-frequency structure and spatial stationarity of certain sources like speech, by using a coherence test to identify low-rank time-frequency bins. The data can be combined coherently or incoherently to arrive at directional spectra which yield the location estimates.

a. Biologically inspired source localization. Of particular interest to this work are approaches taken by bioacoustics and auditory researchers. Two lines of research are notable, namely, those with applications to human hearing and those with applications to animal call monitoring and localization. A biologically-inspired model that exploits time differences in the signals arriving at two sensors was originally proposed by Jeffress (1948) to explain how humans localize sounds using two ears. The model uses a coincidence detection mechanism with a dual delay-line to estimate the direction-of-arrival (DOA) of a sound (with two microphones only the direction of a sound can be estimated in a plane, typically the azimuthal plane, but not the source position). Mathematically, these operations are the same as cross-correlating the signals arriving at the two sensors and determining the time instant of the peak. The peak time is an estimate of the delay between the two sensors, and so the method provides a means for calculating the DOA. The Jeffress model has been validated anatomically and physiologically in the brain of the barn owl (Konishi, 1992) and the cat (Yin and Chan, 1990), and has inspired a number of two-sensor models for estimating DOAs. Most of these models have been applied to auditory processing and have used combinations of time-delays or phase differences, and intensity differences at the sensors to estimate DOAs (Blauert, 1983; Lindemann, 1986; Banks, 1993; Bodden, 1993; Gaik, 1993; Liu *et al.*, 2000).

b. Source localization for bioacoustics monitoring. In field bioacoustics where it is necessary to monitor a population of callers, some of the early multisensor applications were in marine (Watkins and Schevill, 1971; Clark, 1980) and terrestrial (Magyar *et al.*, 1978) domains. However, until recently most applications were confined to localizing and tracking marine mammals (Clark, 1980, 1989; Clark *et al.*, 1986; 1996), and only a few studies have focused on techniques for localization of terrestrial callers[2] (Spiesberger and Fristrup, 1990; Grafe, 1997; McGregor *et al.*, 1997; Mennill *et al.*, 2006). The report of McGregor *et al.* (1997) utilizes the technique due to Clark *et al.* (1996) for locating songbirds. The accuracy of these terrestrial acoustic localization systems, and several of their variants, have been evaluated under both reverberant (Mennill *et al.*, 2006) and non reverberant or moderately reverberant conditions (McGregor *et al.*, 1997; Bower and Clark, 2005). Spiesberger and Fristrup (1990) rigorously derived localization estimates using broadband cross-correlation and Wiener filtering under a variety of signal and environmental conditions. Subsequently Spiesberger developed a technqiue for estimating arrival-time delays from cross-correlation functions that had multiple peaks due to multipath propagation (Spiesberger, 1998), and used it to locate calling birds (Spiesberger, 1999).

The first known application of array localization to frog choruses is the work by Grafe (1997) who used methods due to Clark (1980), Magyar *et al.* (1978), and Spiesberger and Fristrup (1990) to localize a small population of African painted reed frogs. More recently, Simmons *et al.* (2006, 2008) used a small array and a localizer based on a dual-delay line cross-correlator to estimate locations of callers in a bullfrog chorus. To the authors' knowledge, these are the only published reports on localizing callers in a frog chorus. None of the localization methods reported above attempts source recovery, which would be a crucial and necessary technique for detailed analysis of calling behavior.

2. Source recovery

To recover a single sound source with fidelity from a mix of sources, a spatial filter must be designed that passes the target source without distortion but cancels all other competing sources perfectly. The filtering (also known as beamforming) is performed with an array of spatially dispersed sensors. Based on the array data, target to sensor impulse responses (steering vectors) are estimated and a set of filter coefficients are computed. The microphone data are then filtered to yield the required target. Many different beamforming techniques have been developed for a range of applica-

tions (Brandstein and Ward, 2001; van Veen and Buckley, 1988). Filter coefficients may be fixed or adapted to the signal and noise conditions, implementation may be in the time or frequency domain, or the beamformer may actively cancel interfering sources by steering nulls in their directions. We briefly review some of the developments pertinent to this work.

a. Time-domain beamforming. The simplest beamformer is the fixed (i.e., nonadaptive) delay-and-sum beamformer, which for two microphones, is the average of the two microphone signals after one of them is shifted in time to compensate for the intermicrophone delay induced by a target source. This produces on average a 3 dB gain for the target source. However, it is often possible to do much better than a fixed beamformer by adapting the beamformer parameters. There are many techniques and real-time algorithms that are available for adaptive beamforming (Brandstein and Ward, 2001; van Veen and Buckley, 1988). Two commonly used iterative time-domain adaptive beamformers were developed by Frost (1972), and Griffiths and Jim (1982). The relative merits of these processors are discussed in Lockwood *et al.*, 2004, where it is noted that they perform well when canceling interference sources that are statistically stationary and uncorrelated with the target source, but their slow adaptation causes poor performance when confronted with more sources than sensors and when the sources are nonstationary. This is of particular interest as bioacoustic sources in a dense chorus are nonstationary and likely to be more numerous than the sensors. An explicit solution of the optimal beamformer proposed by Capon (1969) avoids the convergence problems of iterative algorithms but for broadband sources requires the inversion of large time-domain correlation matrices. This can be computationally difficult.

b. Frequency-domain beamforming. Implementing a beamformer in the frequency domain eliminates some problems encountered in the time domain. The frequency-domain technique of Liu *et al.* (2000) first localizes sources (see above) and then cancels interfering sources in each frequency band. It can often cancel more sources than sensors. The LENS algorithm (Deslodge, 1998) uses n sensors to actively steer $n-1$ spatial nulls for interference cancellation. More efficient than these algorithms are a class of frequency-domain minimum variance distortionless response (MVDR) beamformers that minimize the energy from interfering sources (minimum variance) while allowing the signal to pass through with unity gain in all frequency bins (distortionless response) (Cox *et al.*, 1986, 1987; Lockwood *et al.*, 2004). Briefly, given the impulse response from the target-source to sensors (the steering vector) and the correlation matrix, an optimum weight vector that specifies how the sensor outputs are to be combined is obtained using the method proposed by Capon (1969). This weight vector is obtained for each frequency bin and applied to the sensor outputs. The weights are computed afresh whenever a new block of data is processed (Lockwood *et al.*, 2004). A variant of this beamformer has also been proposed and applied to small-aperture arrays, and the reported target-source gain ranges from 11 to 14 dB (Lockwood and Jones, 2006). Beamformers using two sensors or a combination of sensors incorporated in two separate packages have found extensive application in the development of binaural hearing aids. Interested readers are referred to Lockwood *et al.* (2004) and Lockwood and Jones (2006) for a review and summary of these algorithms.

C. On-going challenges in bioacoustical monitoring

While localization of vocalizing frogs has been attempted in a few studies (see above), extraction of sources has hardly received any attention. The only attempts in this direction have been highly specialized and have utilized close-mic recording methods to isolate individual callers. For instance, the voice of individual callers has been recovered in artifical ponds using a microphone placed close to a calling perch (Schwartz, 2001), and another study has tracked the voice onset and offset times of individual callers in a natural chorus of coqui frogs (Brush and Narins, 1989). While these techniques are elegant and have provided valuable data on call interactions, they cannot generally be applied to natural choruses. These studies nevertheless provided a motivation for the current work. To take one example, Brush and Narins (1989) were able to determine that a male coqui frog actively avoided call interference with at most two neighbors by adjusting its call timing. Such detailed timing information on each individual caller would be of great benefit to studies on vocal communication if they could be determined on a larger spatial scale. Hitherto it has been difficult to determine the temporal interactions between callers and the spatial distances over which they persist. Thus, determining the spatio-temporal interactions in a chorus requires both source localization and source recovery. The current work is motivated by a need to develop a systematic and overarching framework for tackling this problem. Its development would assist not just the anuran vocalization community, but it would be of interest to other researchers in vocal communication, ethology, ecology, and environmental monitoring.

D. Current work

This report takes a step toward unifying source localization and source recovery into one scheme. It details a processor for blind localization and blind recovery of sources using a microphone array. A large-aperture microphone array is deployed around a frog spawning site. Callers are localized using a gradient-descent approach to solve for the intersection of the hyperboloids resulting from the TOADs. For each localized caller, a steering vector is estimated. This is followed by source recovery using a modified version of the MVDR proposed by Lockwood *et al.* (2004). The result is a spatial and temporal map of the chorus as it evolves in time. Along with theoretical methods, algorithms that detail the link between the various steps are provided. The method is independent of the array size or the number of microphones, and the array can be deployed in any configuration. Tests were carried out using four microphones in a chorus of nine individuals calling in a 9×9 m^2 area. We show that sources can be localized and recovered under non-reverberant or mildly reverberant conditions. The method has not been tested when there is multi-path propagation. We believe that

without improving time-delay estimation, source localization will be difficult under such conditions. This is an important direction for future work.

II. A NOVEL METHOD FOR UNDERDETERMINED BLIND SOURCE LOCALIZATION AND RECOVERY

Consider a chorus with Q distinct species each with L_k, $k=1,\ldots,Q$ individual callers. Each caller is denoted by C_i^k, with $k=1,\ldots,Q$ denoting the species, and $i=1,\ldots,L_k$ denoting the individual within a species. Let $s_i^k(t)$ be the temporal waveform of the caller C_i^k originating from spatial location $\boldsymbol{\zeta}_i^k \in R^3$. The total number of sources L is the sum over all L_k. The caller waveforms $[s(t)]$, spatial locations $(\boldsymbol{\zeta})$, and number of callers (L) are all unknown, but the number of species (Q) is assumed to be known. The following assumptions apply: (i) The spectrum $S_i^k(f)$, considered over all individuals i in species k, is bandlimited to $f \in [f_l^k, f_u^k]$. The lower (f_l^k) and upper (f_u^k) cut-off frequencies are known. They define the frequency band B_k for the species k. (ii) If $s_i^k(n)$ are the samples at times $t=nT$, the data sampling interval is set at a fixed $T=1/f_s$ so that the sampling rate $f_s \geq 2 \max_k\{f_u^k\}$. (iii) The individual source waveforms are independent and block stationary, i.e., over adjacent N-sample intervals, the sources $s_i^k(n)$ are stationary and $E[s_i^k(n)s_j^m(n-l)]=0$, $\forall k,m$, and $\forall i \neq j$ when $k=m$, $|l-n|<N$. (iv) The source locations $\boldsymbol{\zeta}_i^k$ are spatially stationary over the same time scale.

Consider a sensor array with M microphones at known locations $\boldsymbol{\xi}_j \in R^3$ $j=1,\ldots,M$. Let $\boldsymbol{z}(n) \in R^M$ be the discrete-time signal at the array output after sampling at the rate f_s. No assumptions are made about the impulse response between source and sensor other than that it incorporates a time delay dependent on the source-sensor distance. For each species k, the signal $\boldsymbol{z}(n)$ is filtered to pass B_k, resulting in a set of species-specific signals $\boldsymbol{y}^k(n) \in R^M$, $k=1,\ldots,Q$.

Given the species-specific signals $\boldsymbol{y}^k(n)$ and microphone locations $\boldsymbol{\xi}$, the goal is to estimate the location $\boldsymbol{\zeta}_i^k$ and the call $s_i^k(n)$ for each individual i belonging to species k.

This section is divided as follows. Theoretical methods for locating a source are presented in Sec. II A and those for adaptive beamforming are presented in Sec. II B. The implementation requires sources to be localized first, and then recovered using the beamformer. The algorithms and analysis that integrate localizing and beamforming are detailed in Sec. II C.

A. Acoustic source localization

Let vector $\boldsymbol{\xi}_j$ denote the spatial coordinate (x,y,z) of sensor j with reference to an aribitrary origin. Let the source coordinate be denoted by $\boldsymbol{\zeta}$. The source-sensor distance is defined as $D_j=\|\boldsymbol{\xi}_j-\boldsymbol{\zeta}\|$. Only one source is considered because the processing scheme considered below selects only those time-frames where a single source is dominant. The range difference between two sensors i and j will be denoted by

$$d_{ij}=D_i-D_j, \quad i,j=1,\ldots,M. \tag{1}$$

Let the vector of all possible range differences be denoted by $\boldsymbol{d}$. There will be $M(M-1)/2$ distinct range-difference measurements for all microphone pairs. For localizing sources in three dimensions the system is overdetermined whenever $M \geq 4$ or $M \geq 5$ depending on the source sensor geometry (see footnote 1). In this case the redundant measurements can provide improved estimates in noise. The geometrical relationship between the source and sensors are depicted in Fig. 1.

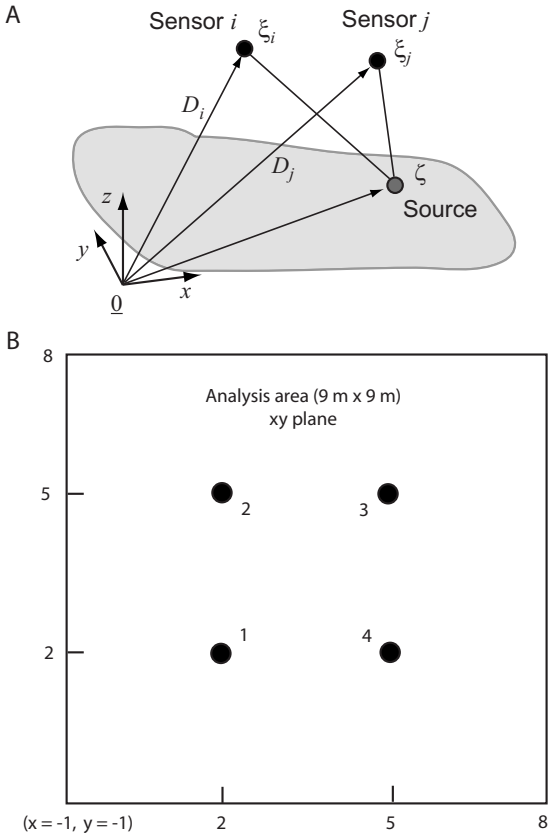


FIG. 1. Source and sensor geometry. (A) Shown is a single source located at $\boldsymbol{\zeta}$ and two sensors i and j at locations $\boldsymbol{\xi}_i$ and $\boldsymbol{\xi}_j$, respectively. Directed arrows represent vectors from an arbitrary origin 0. Source-to-sensor distances are marked as D_i and D_j. The arrival-time differences are $\tau_{ij}=c^{-1}(D_i-D_j)$, where c is the velocity of sound, or equivalently the range differences $d_{ij}=D_i-D_j$. (B) Deployment of sensors in the field test, looking down on to ground (along z-axis). Ground is xy plane with $z=0$. Omnidirectional microphones numbered 1–4 (filled circles) were placed in a square with the following (x,y,z) coordinates. (1) (2.0, 2.0, 1.69), (2) (2.0, 5.0, 1.74), (3) (5.0, 5.0, 1.76), and (4) (5.0, 2.0, 1.65). The analysis area extended to a square 9 m on either side (solid line, not to scale) with the lower-left corner at $x=-1$, $y=-1$.

The source location $\boldsymbol{\zeta}$ is to be determined given the vector of range-difference measurements $\boldsymbol{d}$ and the microphone locations $\boldsymbol{\xi}$. An iterative gradient-descent optimization technique is followed. Given an estimate $\hat{\boldsymbol{\zeta}}$ of the unknown source location, the range-difference estimate $\hat{\boldsymbol{d}}=\|\boldsymbol{\xi}-\hat{\boldsymbol{\zeta}}\|$ is formed. The error in the delay $\boldsymbol{\epsilon}=\mathbf{d}-\hat{\mathbf{d}}$ is determined and the goal is to minimize the squared delay error, i.e.,

$$\min\{\boldsymbol{\epsilon}^T\boldsymbol{\epsilon}\}. \tag{2}$$

For a fixed location step $\Delta\boldsymbol{\zeta}$ the gradient of the squared error is calculated, and the new estimate of $\hat{\boldsymbol{\zeta}}$ is determined from

$$\hat{\boldsymbol{\zeta}} \leftarrow \hat{\boldsymbol{\zeta}} - \mu \frac{\Delta(\boldsymbol{\epsilon}^T \boldsymbol{\epsilon})}{\Delta \boldsymbol{\zeta}}. \tag{3}$$

The parameter μ is adaptively adjusted so that the squared error $\boldsymbol{\epsilon}^T \boldsymbol{\epsilon}$ is made smaller. The initial value of $\hat{\boldsymbol{\zeta}}$ is taken to be some suitable value, say, the mean value of $\boldsymbol{\xi}$. An approximate solution to the problem of intersecting hyperboloids based on spherical interpolation has been proposed by Smith and Abel (1987). The authors suggest that their solution could be used as an initial condition in iterative nonlinear minimization methods, such as the gradient-descent method proposed here, so as to improve convergence. But this has not been attempted here.

B. Acoustic beamforming

Recovery of the individual vocalizations by beamforming requires a steering vector, which at each frequency defines the relative amplitude and phase relationships of a signal from the location of interest for all microphones in an array. A MVDR beamformer preserves any signal exhibiting the exact amplitude and phase relationships defined by the normalized steering vector (distortionless response) while minimizing the interference energy in the output (minimum variance) from sources with any other amplitude and phase relationships between sensors (Capon, 1969). This precision potentially allows the separation of vocalizations of even closely spaced frogs.

Let the microphone signals be denoted by $\boldsymbol{y}(n) \in R^M$. Data frames of length N_F are windowed and Fourier-transformed via an N-point FFT (Fast Fourier Transform). The frequency-domain signals are denoted by $\boldsymbol{Y}(f)$ where f represents frequency. The cross-correlation matrix $\boldsymbol{R}(f)$ at frequency f (an $M \times M$ matrix) is computed from $\boldsymbol{E}[\boldsymbol{Y}\boldsymbol{Y}^H(f)]$, where $\boldsymbol{Y}^H$ represents the transposed complex conjugate (Hermitian) of $\boldsymbol{Y}$. For the remainder of the discussion the frequency f will be omitted for notational simplicity and it will be assumed that the quantities are confined to a frequency bin unless otherwise stated. If there is only a single source i that is present in the data frame, then the cross-correlation matrix consists of the outer product of the steering vector $\boldsymbol{e}_i$ times the power of the signal (σ_s^2),

$$\boldsymbol{R} = \sigma_s^2 \boldsymbol{e}_i \boldsymbol{e}_i^H. \tag{4}$$

The cross-correlation matrix $\boldsymbol{R}$ from each data frame is computed. Because only one source dominates the microphone data, $\boldsymbol{R}$ is rank-1 with an eigenvector that is an energy-normalized version of $\boldsymbol{e}_i$. The eigenvector $\tilde{\boldsymbol{e}}_i$ corresponding to the largest eigenvalue of $\boldsymbol{R}$ is an estimate of the steering vector for that source at each frequency. The estimated steering vector is scaled so that the steering-vector element at some selected microphone m is 1, i.e.,

$$\boldsymbol{e}_i = \bar{\boldsymbol{e}}_i e_{m,i}. \tag{5}$$

With known steering vectors $\boldsymbol{e}_i$ to source i, the optimal weights ($\boldsymbol{w}_i^*$) for combining the different microphone channels are obtained from the solution to a linearly constrained quadratic optimization problem. The MVDR beamformer minimizes the output power subject to the constraint that the gain of the target signal is unity, i.e.,

$$\min_{e_i^H w_i = 1} \boldsymbol{E}[|\boldsymbol{w}_i^H \boldsymbol{Y}|^2]. \tag{6}$$

In this case the optimal weights are (Capon, 1969)

$$\boldsymbol{w}_i^* = \frac{\boldsymbol{R}^{-1}\boldsymbol{e}_i}{\boldsymbol{e}_i^H \boldsymbol{R}^{-1} \boldsymbol{e}_i}. \tag{7}$$

The optimal frequency weight $\boldsymbol{w}_i^*$ is applied to each frequency bin (of the Fourier-transformed data) at each time frame over which the correlation matrix $\boldsymbol{R}$ was computed. Then an inverse Fourier transform is performed to recover the source. The time-frames are then pieced back together to recover the entire vocalization from the given source for all times. By determining the steering vector and optimum weights for each source location, all individual callers can be extracted with minimum distortion.

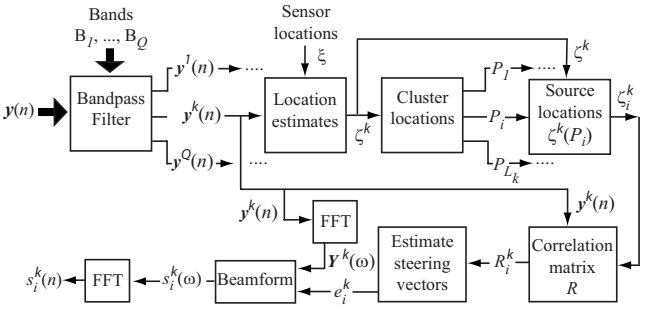


FIG. 2. The source localization and source recovery processor. The procedure is outlined for localization and recovery of a single source i from a single species k. These quantities are $\boldsymbol{\zeta}_i^k$ and $s_i^k(n)$, respectively. Notation follows text.

C. Algorithms and analysis procedure

Acoustic data from a chorus are recorded synchronously at multiple microphones positioned around a spawning site. Subsequently, using an offline procedure, the location of each individual caller in the neighborhood of the microphone array is estimated using a source localizer (Sec. II A), and its call recovered using an acoustic beamformer (Sec. II B). The localization and beamformer steps are linked in a six-step procedure, and schematically reproduced in Fig. 2.

(1) *Bandpass filter raw microphone data*. For each species k, microphone data $\boldsymbol{z}(n)$ are bandpass filtered into bands B_k resulting in microphone data sets $\boldsymbol{y}^k(n)$. Bandpass filtering retains the spectral range of that species (the "focal species") while minimizing interference from other species. Even if there is some overlap with the frequency band of other species, isolating the bandwidth of the focal species leads to greater localization accuracy, as it reduces interference and improves beamformer performance. Subsequent steps are applied individually to each band B_k; i.e., analysis of a focal species is independent of other species present in the chorus.

(2) *Find time intervals with single dominant call*. Experimental data indicate that even in a dense chorus, many

time intervals contain only a single strong call within the spectral range of a given species from an individual within or near the sensor array. From these times the location of the single caller can be determined by cross-correlating the signals at the various microphone pairs (i,j) and finding the relative time-delay (τ_{ij}^*) of maximum correlation between each pair of sensors. Let $y_i^k(n)$ and $y_j^k(n)$ be the bandpass-filtered outputs for species k from microphones i and j, respectively. The correlation coefficient of the cross-correlation at the maximizing delay, τ_{ij}^*, is

$$\rho_{ij} = \frac{\Sigma_{r=l}^{n} y_i^k(r) y_j^k(r+\tau_{ij}^*)}{\sqrt{\Sigma_{r=l}^{n}(y_i^k(r))^2}\sqrt{\Sigma_{r=l}^{n}(y_j^k(r+\tau_{ij}^*))^2}}, \tag{8}$$

where l and n are the start and end times of the interval of analysis, and the correlation frame length is $n-l+1$. If the waveforms from the different microphones are identical replicas (within a scale factor and a time-delay), then the correlation coefficient ρ_{ij} at the maximizing delay τ_{ij}^* will be 1. If there are multiple sources of similar power originating from different locations, ρ_{ij} will be considerably smaller. Thus, we apply a threshold γ to all pairwise coefficients so that whenever $\rho_{ij} \geq \gamma$ the interval is considered to be dominated by a single source. The value of γ is determined experimentally. The set of frames for which $\rho_{ij} \geq \gamma$ will be denoted by P. For a given frame in P, the maximizing delay τ_{ij}^* for every pair of microphones is denoted by the vector $\boldsymbol{\tau}^*$. For all frames in P this step yields a set of $\boldsymbol{\tau}^*$ which can be used to estimate the source location. In certain situations it may be advantageous to consider pairwise correlations from subsets of sensors instead of all sensors, for instance, when some sources are close to a subset of sensors but much farther away from the remaining sensors.

(3) *Find location of dominant source intervals by least-squared delay error*. For any given $\boldsymbol{\tau}^*$ the measured range differences are given by $\boldsymbol{d}=c\boldsymbol{\tau}^*$, where c is the velocity of sound. The location of the source corresponding to the range difference is obtained via the gradient-descent procedure outlined in Sec. II A. The localization is performed on every frame in P and results in a set of location estimates, each corresponding to a single caller. The identities of the callers C_i^k are, however, still unknown.

(4) *Cluster the location estimates to identify individual frogs*. If the data record $\boldsymbol{y}^k$ is sufficiently long, then each caller is likely to dominate one or more frames with indices P_i such that $\boldsymbol{\zeta}^k(P_i)$ corresponds to a single caller C_i^k. These location estimates will be somewhat variable due to measurement noise even though the individual is stationary. However, it is assumed that the variability in the estimates is smaller than the mean spacing between frogs, thus providing a natural way to cluster the location estimates $\boldsymbol{\zeta}^k$ into sets $\boldsymbol{\zeta}^k(P_i)$ each corresponding to a caller C_i^k. From this cluster, the averaged spatial location $\boldsymbol{\xi}_i^k$ of each actively calling frog can be determined. Here the clusters are determined visually from localization plots, although automatic clustering procedures can be applied to extract the sets $\boldsymbol{\zeta}^k(P_i)$.

(5) *Estimate beamformer steering vectors*. Small deviations in the steering vectors can cause the adaptive beamformer to treat the target source as interference and to cancel it as well. In frog choruses variations in the steering vector can result from (1) minor deviations in microphone placement, (2) any reflections or absorption of sound from the ground, (3) presence of other objects such as vegetation in the environment, or (4) direction-dependent acoustic radiation patterns in vocalizations. These can make recovery of calls difficult if not impossible with current methods. The following new procedure has proven effective in blindly estimating the steering vectors from field recordings with high accuracy.
The cross-correlation matrix $\boldsymbol{R}$ [see Eq. (4)] is computed for the frames $\boldsymbol{y}^k(P_i)$ and averaged in each frequency band. Recall that this set of frames corresponds to the cluster of location estimates $\boldsymbol{\zeta}^k(P_i)$ for caller C_i^k. Because only one caller C_i^k dominates the cluster, the rank of $\boldsymbol{R}$ is essentially 1 and the eigenvector corresponding to the largest eigenvalue of $\boldsymbol{R}$ is an estimate of the steering vector $\boldsymbol{e}_i^k$ for the caller. The steering vectors are renormalized across all frequencies according to Eq. (5) to reconstruct the frog vocalization without distortion at the closest microphone. The procedure is repeated to estimate the steering vectors for each individual in the chorus.

(6) *Beamform to recover individual acoustic signal at all times*. For each caller C_i^k the optimal weights $(\boldsymbol{w}_i^{k*})$ for combining the different microphone channels of the MVDR adaptive beamformer are computed in each frequency bin according to Eq. (7) and the beamformer output is calculated using

$$S_i^k(f) = \boldsymbol{w}_i^{k*}(f)^H \boldsymbol{Y}^k(f). \tag{9}$$

The vocalization $s_i^k(n)$ of caller C_i^k is obtained via the inverse Fourier-transform of $S_i^k(f)$. The procedure is repeatedly applied to recover all sources.

III. FIELD TESTING

Recordings of choruses were carried out at a spring-fed marsh located in the Cibolo Nature Center (Boerne, TX). The site coordinates were 29°47′7.51″ N, 98°42′37.92″ W at an elevation of 422 m. A 7×7 m^2 grid (1 m spacing) was marked using short stakes driven into the ground. The grid was used for visually locating calling individuals, but the analysis was carried out over an 81 sq m area (9 m to a side). Four omnidirectional microphones numbered 1–4 (Sennheiser MKE-2, 0.02–20 kHz) were positioned at the vertices of a square 3 m on each side centered in the grid [Fig. 1(B)]. All microphones were mounted on poles. In this coordinate system $\boldsymbol{\xi}$ represents (x,y,z) with the z axis being normal to the ground (where $z=0$). Microphone data from the array (number of sensors, $M=4$) were used for source local-

ization and source separation. Note that the ground was planar ($z=0$); therefore each source could be located precisely with four microphones (see footnote 1).

The omnidirectional microphone outputs were amplified using battery-powered field amplifiers (Sound Devices MP-1) and the cables from the array were led into a blind that housed the data-acquisition system and other components. Microphone data were acquired synchronously at a sampling rate of 20 kHz (National Instruments PXI 4472, eight-channel, 24 bit) by a data-acquisition computer (National Instruments PXI-8186 controller running Windows XP, mounted in a PXI 1031DC chassis). Data-acquisition programs were developed in the LabVIEW environment (National Instruments Inc.). All equipment were powered using DC (battery) sources. Recorded data were analyzed offline using MATLAB (The MathWorks Inc.)

Once the chorus was in progress the frogs and toads that could be visually located within the grid area were identified, and their x and y locations were marked on a chart (the z coordinate was assumed to be 0, as all the located individuals were calling from the ground). The uncertainty of the visual estimates were estimated to be about 10 cm along x and y directions. The area was then vacated and microphone data were acquired for a period ranging from 5 to 10 min. Then the positions of the previously marked frogs were once again determined, and new entrants if any were noted. This procedure was repeated for the duration of the session. At the end of the session only the microphones were removed, but the grid and microphone stands were left intact at the site. This ensured that microphone locations between sessions were unchanged.

IV. RESULTS

Recordings were carried out in mid-March 2007 between 2100 and 2400 h. Two species were present on all days in this period. They were *Bufo valliceps* (Gulf Coast toad) and *Acris crepitans blanchardi* (Blanchard's cricket frog), abbreviated *Bv* and *Ac*, respectively. Cricket frogs outnumbered Gulf Coast toads in the entire site. The Rio Grande leopard frog *Rana berlandieri* was present from late February to early March. It was identified visually and by voice but was not present during the days the chorus was recorded. Data presented here are for a mixed *Bv* and *Ac* chorus, and were collected on March 21, 2007. Average temperature was 21 °C, relative humidity was 78%, and pressure was 1019.5 hPa.

For the duration of the recording, two *Bv* males were identified visually in the 9×9 m^2 arena [*Bv*1: (2.6, 4.5), *Bv*2: (2.0, 4.5)]. Cricket frogs were harder to identify due to their small size and coloration, and only one was visually located [*Ac*1: (1.95, 4.6)]. Although not all callers within the arena were visually located, as will be seen in the results below, the known locations of *Bv*1, *Bv*2, and *Ac*1 will be compared to the computed locations so as to verify the accuracy of the localization algorithm.

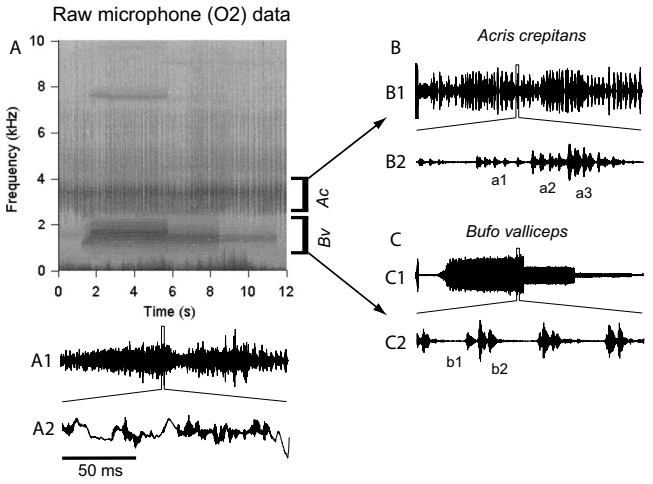


FIG. 3. Bandpass filtering of raw microphone data into species-specific bands. (A) Spectrogram of mic 2 output. Two distinct species-specific bands marked *Bv* and *Ac* can be discerned. Panel A1 shows the time waveform of the signal. A 150 ms window from the segment is expanded and shown in detail in panel A2. (B) depicts the bandpass-filtered waveform corresponding to band *Ac* for the cricket frog (panel B1), with fine temporal details corresponding to the 150 ms window shown in panel B2. Three individuals marked a1, a2, and a3 are calling in this window. (C) depicts similar results for the Gulf Coast toad filtered into band *Bv*. The window depicts two callers b1 and b2, with b2 being more intense than b1.

A. Call characteristics

Figure 3(A) depicts the spectrogram for a 12 s segment of the chorus recorded at microphone 2. The dominant call frequencies of *Bv* (species 1) and *Ac* (species 2) were spectrally separated into non-overlapping frequency bands: 680–2300 Hz (spectral band B_1) and 2700–4000 Hz (spectral band B_2), respectively. Several individuals from both species were calling in this segment, and there was extensive temporal overlap within and across species. Calls of one *Bv* individual contained some harmonic components, most notably the fifth harmonic of the dominant frequency [between 7 and 8 kHz; see call spectrogram Fig. 3(A), 1.5–5.5 s]. Power in this band was attenuated by 32 dB with respect to B_1. Calls of *Ac* also possessed harmonics but because of their pulsatile nature, energy was distributed over a broad range of frequencies (between 4.5 and 7 kHz). Power in this band was attenuated by 20 dB with respect to B_2.

In general, harmonic components were greatly attenuated with respect to the bands B_k and so it was assumed that neglecting the higher harmonics would not make a significant difference to either call localization or separation. Further, the harmonics could get washed out in background noise depending on the proximity and orientation of the caller with respect to the microphones. The *Bv* individual mentioned above was close to mic 2 and positioned so that the fifth harmonic was captured, whereas those of the other toads calling at the same time (see panel C) were not distinguishable from the background. Thus the use of higher harmonics may not provide additional benefit and could, in fact, degrade localization and beamformer performance by reducing signal-to-noise ratio. Small changes in the upper and lower cut-off frequencies of the bands did not significantly affect processing. Energy in the band below 500 Hz was primarily from wind and other abiotic sources. Microphone data

corresponding to the spectrogram are shown in Fig. 3(A1). Panel A2 shows waveform details of a 150 ms segment selected from Panel A1 (windowed portion). The high-frequency fluctuations (dark bands) are *Bv* and *Ac* callers. The slow fluctuations are due to wind and other low-frequency noise sources.

The spectrogram demonstrates that the two species are reasonably well separated in frequency space and that their calls can be processed independently of one another by bandpass filtering the data into two parallel data streams. Accordingly, data from each microphone were filtered into two streams $\boldsymbol{y}^1$ (*Bv*) and $\boldsymbol{y}^2$ (*Ac*) restricted to bands B_1 and B_2, respectively. The filter outputs for the mixed waveforms shown in [Fig. 3(A1)] are depicted in (Figs. 3(B) and 3(C)), for the bands B_1 and B_2, respectively. The fine temporal details corresponding to the 150 ms segment (panel A2) are shown in panels B2 and C2. Multiple callers within each species can be discerned. There are three *Ac* individuals (a1, a2, and a3), and two *Bv* individuals (b1 and b2) distinguishable by their relative amplitudes. Leakage of calls from one species into the band of the other was insignificant. Broadly, panels B and C demonstrate that calls can be unmixed into species-specific streams. The results shown in Figs. 3(B) and 3(C) are analogous to the spectral filtering that takes place in the inner ears of anurans (Capranica, 1965).

B. Localization of callers

Hereafter the same procedure was applied to data from both species. Location was computed on a frame-by-frame basis using a correlation block size of 500 ms (*Bv*) and 20 ms (*Ac*). For each frame the pairwise normalized cross-correlation function was computed and its maximum value ρ was determined. A threshold γ=0.65 was apllied to ρ. When ρ was greater than γ the frame was assumed to contain only a single caller and was retained; otherwise, it was discarded. Figure 4 illustrates the process. To illustrate the procedure, the bandpass-filtered traces from microphones 1 and 2 are shown in Fig. 4(A) [filterband B_1 corresponding to *Bv*, identical to the segment shown in Fig. 3(C1)]. Three representative time frames at which the cross-correlation functions were evaluated are shown in panel A as rectangles marked B, C, and D. Cross-correlation functions in each of these frames are shown in the respective panels. The selected frames show two toads 1 and 2. In frame B only toad 1 is present [Fig. 4(B): ρ=0.97, τ^*_{ij}=2.8 ms], in frame D only toad 2 is present [Fig. 4(D): ρ=0.97, τ^*_{ij}=−5.1 ms], whereas in frame C both toads were present [Fig. 4(C): ρ=0.42, and ρ=0.51 ms at the maximizing delays shown in Figs. 4(B) and 4(C), respectively]. After applying the threshold only frames B and D were retained for estimating steering vectors because they contained only one dominant caller, whereas frame C was discarded.

In a 1 min interval approximately 3% of the frames for cricket frogs were identified as having one dominant caller (totaling at least six callers). For Gulf Coast toads the number of frames with a single caller was approximately 34% (totalling four callers). All calling individuals were represented at least once in the selected frames as most of them

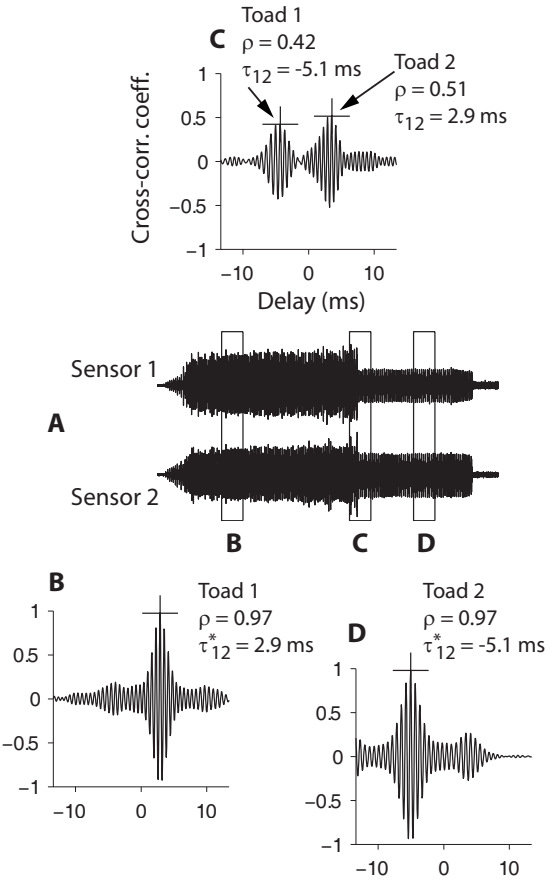


FIG. 4. Time-domain cross-correlation to test for the presence of a single caller in a selected frame. (A) shows bandpass (B_1) filtered outputs from two sensors 1 and 2. Three representative time frames B, C, D are shown as rectangular windows. The corresponding cross-correlation functions are shown in panels B, C, and D, respectively. For frames B and D, the maximum in the cross-correlation function is marked (+), and the ρ and maximizing delay τ^*_{12} are also shown. These frames have one dominant caller each: toad 1 (frame B) and toad 2 (frame D) and ρ exceeds threshold (γ =0▷65). Frame C has both callers as seen from the two peaks in the cross-correlation function at the delays exhibited by toad 1 (panel B) and toad 2 (panel D). Consequently, the cross-correlation function is broad and neither peak exceeds γ. This frame was discarded. Note that the two callers can be visibly distinguished in the sensor data except in the overlapping region (see also the identical segment shown in Figs. 3(B1) and 3(B2)).

positioned their calls to avoid overlap. For each frame the maximizing delays $\boldsymbol{\tau}^*$ were converted to a range-difference estimate $\boldsymbol{d}=c\boldsymbol{\tau}^*$ [see Eq. (1) and Step 3 in Sec. II C]. This set of frames is denoted by P.

After analyzing all frames and determining the vector $\boldsymbol{d}(P)$, the source-localization procedure outlined in Sec. II A was applied to each frame in P and a raw position estimate was obtained for that source. Position estimates of single callers $\boldsymbol{\zeta}^k(P)$ (k=1: *Bv*; k=2, *Ac*) were graphically plotted. The identities of the callers at this point were still unknown because the location estimates differed across individuals (due to spatial separation). However, estimates for a single individual were also subject to variability due to measurement noise. Thus the frames $P_i \in P$ which correspond to individual i had to be determined by a clustering procedure. This visual procedure relied on the small variability in an individual's location in comparison with the inter-individual spacing.

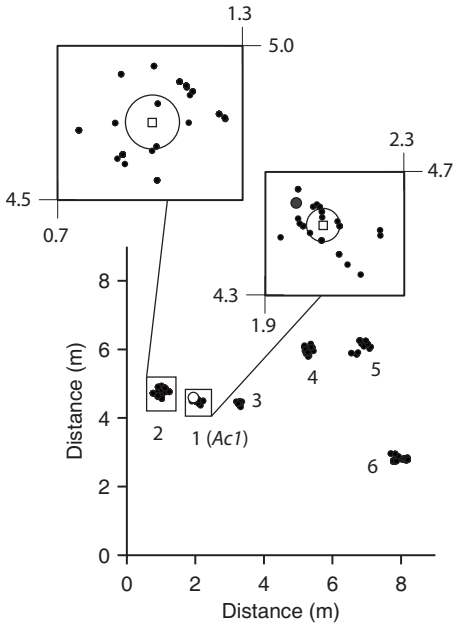


FIG. 5. Two-dimensional map of cricket-frog locations. Frames of 20 ms duration containing only one caller were identified by cross-correlation, and the location of the calling individual was determined within the 9 m−9 m area. Point estimates were clustered visually, and six frogs (numbered 1–6) were estimated to be present in the arena. Frog 1 (*Ac*1) was visually located prior to the recording at the position marked with an open circle. Insets show the clustering procedure for frogs 1 (*Ac*1) and 2. The point estimates were clustered visually by selecting the bottom-left and top-right corners of a bounding box. Points within the box were assigned to a single caller with mean position indicated by a square. The circle defines the positions within one standard deviation of the mean position.

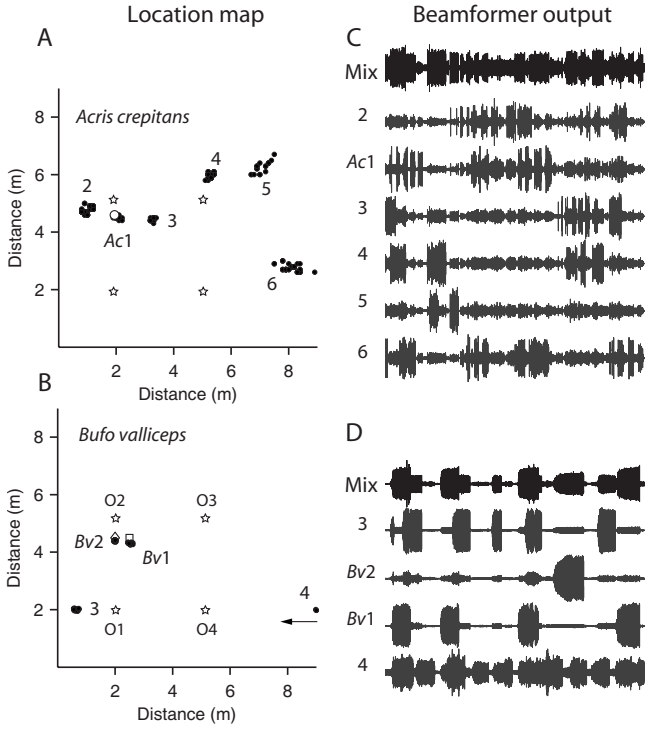


FIG. 6. Location maps [(A) and (B)] for cricket frogs and Gulf Coast toads in the *xy* plane (ground), and their beamformer outputs [(C) and (D)]. Microphone locations are shown as 1–4 [open stars in (A) and (B)]. The location of each individual is a numbered cluster. Visually observed individuals are marked *Ac*1 (panel A, ○) and *Bv*1 and *Bv*2 (panel B, ◇ and □, respectively). In panel B cluster 4 represents two toads located outside the analysis arena ($x>9$ m, $y<2$ m). See text for further explanation. Beamformer outputs for each individual are shown in gray. The filtered bandpass output from a representative microphone (mic 2) are shown in black ("Mix").

This is illustrated in Fig. 5 for cricket frogs. Based on a visual clustering procedure, six cricket frogs were identified. Their estimated mean positions are listed in Table I. The individual who had been visually located (frog 1 or *Ac*1) is listed and his position marked on the graph in Fig. 5 (open circle). An identical procedure was applied to the spectral band B_2 to estimate the locations of Gulf Coast toads. The results are listed in Table I along with the locations from visual estimates for *Bv*1 and *Bv*2. For both species, the location estimates from the data were in good agreement with the visual estimates where available. Location maps for both species are shown in Figs. 6(A) and 6(B).

C. Beamforming and call separation

The clustering procedure identified the data frames P_i for each individual. From these frames the mean location ζ_i^k

TABLE I. Coordinates (x,y) of individual callers in meters ($z=0$ for all individuals). "Algorithm": locations calculated from microphone data. "Visual": locations determined visually for some individuals. Standard deviations are indicated below the coordinates. No. 4 under *B. valliceps* was a cluster of two toads with the *x* and *y* positions beyond the range of the algorithm. See also location maps in Figs. 6(A) and 6(B).

	B. valliceps (x,y)		*A. c. blanchardi* (x,y)	
No.	Algorithm	Visual	Algorithm	Visual
(1)	(−0.51, 1.17)		(2.03, 4.53)	Ac1: (1.95, 4.6)
	(±0.031, ±0.024)		(±0.064, ±0.067)	
(2)	(2.03, 4.37)	Bv2: (2.0, 4.5)	(1.0, 4.75)	
	(±0.003, ±0.003)		(±0.145, ±0.108)	
(3)	(2.55, 4.27)	Bv1: (2.6, 4.5)	(3.28, 4.43)	
	(±0.006, ±0.004)		(±0.036, ±0.04)	
(4)	(>9, <2.0)		(5.26, 5.96)	
	2 toad cluster		(±0.051, ±0.067)	
(5)			(6.9, 6.16)	
			(±0.217, ±0.184)	
(6)			(8.01, 2.77)	
			(±0.195, ±0.076)	

and the mean correlation matrix $\boldsymbol{R}_i$ was computed, and the steering vector was estimated following the proceedure outlined in Step 5 in Secs. II C and II B. The minimum-variance beamformer used the estimated steering vector $\boldsymbol{e}_i$ for each individual to recover its calls for all times, while suppressing all other interfering sources. For each recovered source, the beamformer output is a single channel, one per individual, available at the same sampling rate as the raw data (20 kHz). Traces of the recovered sources are shown in Figs. 6(C) and 6(D), and numbered according to the source depicted in Figs. 6(A) and 6(B), respectively. The *Bv* source marked with an arrow (cluster 4, in panel B) was a cluster of two toads located at $x>9$ and $y<2$. This cluster could not be localized due to the large range. That this cluster has more than one individual can be readily seen from the beamformer output in which the call density is higher than the density for toad 3, *Bv*1, and *Bv*2. Some degree of cross-talk between channels is visible in some of the recovered source channels. For example, the *Bv*1 and toad 3 channels crossover to the *Bv*2 channel.

There is no general characterization (such as a beampattern) of adaptive beamfomer performance, as it depends on the exact array and source configurations and their individual spectra and calling times. But an empirical estimate of the performance can be arrived at by simulations. Sources corresponding to two of the Gulf Coast toads (*Bv*1 and *Bv*2) were synthetically presented to the same array. The toad *Bv*1 was placed at coordinates (3.0, 2.67, 0), and *Bv*2 was located at random on a circle around *Bv*1. Twenty-nine circles with radii spaced logarithmically from 0.02 to 5 m were selected, and around each circle 30 random locations were determined for a total of 870 locations. Figure 7(B) shows the caller *Bv*1 (◇) and all the locations of *Bv*2 tested in the simulations (●). For each pair of *Bv*1 and *Bv*2 locations, the procedure used for localizing and extracting the sources as outlined above was followed. Each source was extracted in turn while suppressing the other.

Let E_1 and E_2 be the energies of the *Bv*1 and *Bv*2 calls, respectively, that were selected for mixing. Following recovery let $\hat{E}_{ij}$ with $i,j=1,2$ be the energy of the source j in target channel i. Attenuations were calculated in dB as $a_{ij}=10\log_{10}(\hat{E}_{ij}/E_i)$. The term a_{12} represents the amount of residual energy (cross-talk) from *Bv*2 in the recovered *Bv*1 channel. The term a_{21} represents the cross-talk resulting from *Bv*1 in the recovered *Bv*2 channel. These should be large and negative. The terms a_{11} and a_{22} represent self-cancellation and should ideally be close to 0 dB. The results are shown in Fig. 7(A). The averages of a_{12} and a_{21} (over 30 locations at each distance) are shown with standard deviations (thin lines). In general there is a sharp decline in cross-talk as the sources are spatially separated up to a critical distance of about 10 cm. At this separation the attenuation is about 45 dB. For progressively larger separations the attenuation gradually decreases to about 25 dB due to increasing relative time-delay differences between the sources. The average self-cancellation of the target source (○) is almost 0 dB over the range of distances indicating that the recovery does not appreciably subtract the desired target.

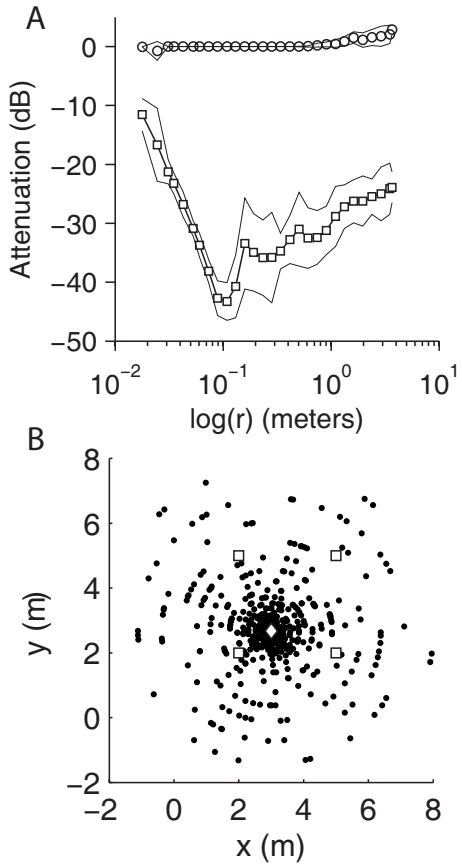


FIG. 7. Beamformer performance as a function of inter-source distance. Calls from two Gulf Coast toads were synthetically combined so that they appeared to originate from distinct locations. Each source was localized and then extracted with the beamformer while the second source was treated as an interference. (A) The average attenuation (dB, ordinate) of the interfering source in the target source (□) is shown as a function of the logarithm of distance (abcissa). Thin lines represent the standard deviations of the attenuation over repeated trials. Also shown are self-cancellation (○), the degree to which a target is canceled by the beamformer. (B) The results in (A) were obtained by fixing one of the toads (◇) at (3.0, 2.67, 0) while the second toad had a variable location (●). The plot shows all the 870 locations (●) that were used to generate the plots in (A). Microphone locations (□) as in Fig. 1(B). See text for details.

V. DISCUSSION

This report details a passive microphone array technique for locating and recovering the calls of vocalizing frogs in a natural chorus. The technique is blind in that it makes no specific assumptions about the sources (callers). Instead, it utilizes heuristics that stem from biological plausibility, in particular, time-frequency sparseness (Mohan *et al.*, 2008). At the core of the processor are two powerful theoretical methods originating from array signal processing: (1) Localization of a source using the TOADs between pairs of microphones and (2) recovering a source by adaptive beamforming. The two parts (localization and recovery) are linked by a novel, yet simple procedure for estimating steering vectors from the location information and then actively steering the array toward the source for recovery. The procedures are repeatedly applied to every source. The end result is a spatiotemporal map of the chorus. The core methods—localization and source recovery using beamforming—have been widely investigated (see Sec. I). What is new in this report is the

processing system, and, in particular, the steering-vector estimation. They are motivated by biological mechanisms for listening in noise. The method will not work without modifications when there are significant multi-path reflections.

A. Segregating data into species-specific spectral bands

The first step takes a set of raw sensor data and bandpass filters it into species-specific bands. The segregated streams of data, one for each species, are subsequently processed in parallel as they are independent of one another. There is good reason to follow this strategy although it increases the computational effort. If the assumption is that conspecific callers avoid overlap, as has been suggested by many studies (see Bee and Micheyl, 2008; Feng and Ratnam, 2000; Gerhardt and Huber, 2002, for reviews), then removing potentially overlapping heterospecifics will make it more likely to find temporally segregated callers. Location estimates become more accurate as there is less energy from spatially separated interfering sources. The step is no different from the matched filtering that takes place in the frog ear originally proposed by Capranica (1965). Namely, the inner ear acts a spectral filter matched to the species-specific mating call, thereby selectively enhancing the calls of its own species while suppressing calls from other species. It is possible to bandpass filter the data into a spectral band that exactly matches the data but it has not been attempted here. In the case of partial spectral overlap between species, location and individual calling times can easily be determined by limiting the recovery to only the nonoverlapping portions of the spectral band. One concern about this step is that it requires *a priori* information on the number of species. For most recordings and locations, this does not present a significant problem as the information is easily obtained by listening to the recorded microphone data.

B. Source localization

To localize a calling individual and estimate the steering vector to his location, it is necessary to find at least one data frame where the frog is the only individual vocalizing near a group of at least four to five microphones in his spectral band. This is possible when data are collected for long durations, giving each individual the opportunity to find temporal gaps when he can vocalize without interference. For instance, in the case of cricket frogs the percentage of data frames where only one caller was present was about 3%, whereas for Gulf Coast toads it is about 34%. The difference in numbers is a result of the call duration and rate. Cricket frogs produce brief pulsatile call notes of about 30 ms duration with a low duty cycle, whereas Gulf Coast toads have a call duration that is about 5 s. Individuals from both species appear to avoid overlap with conspecific callers. This is a general feature of communication in choruses and highlights a common strategy for hearing in noisy environments (see Feng and Ratnam, 2000; Gerhardt and Huber, 2002; Bee and Micheyl, 2008).

The localizer presented here also exploits the biological strategy of "listening in the gaps" to accurately estimate the location of a single dominant caller. It uses a time-domain cross-correlator that selects only those time frames where a single frog is vocalizing [e.g., Figs. 4(B) and 4(D)] while ignoring frames with more than one caller [Fig. 4(C)]. Stated another way, for single-caller frames the sensor covariance matrix $\boldsymbol{R}$ has unity rank. This strategy can be used in choruses that are more dense than the chorus studied here, as the following argument demonstrates.

Typically callers tend to attend to interfering callers in a local neighborhood while callers further away are ignored or remain unnoticed because they are greatly attenuated. Thus we can increase the likelihood of finding single-caller frames by restricting the analyses to data from subsets of sensors that are close together. Reducing the number of sensors in this way would reduce the spatial extent or coverage of the sensors, and restrict the chorus area to a neighborhood in the vicinity of the selected sensors (casting a "spotlight" on the neighborhood). The total energy from interfering sources that are further away is reduced. In these restricted neighborhoods, there is a greater likelihood of finding time windows where the call from a single individual dominates all other callers, and the covariance matrix from the selected subset of sensors will effectively have unity rank. Thus, dense choruses or arrays of large size can be analyzed by selectively restricting the array geometry during post-processing. For these reasons the array geometry (sensor placement) is not critical provided that the number of sensors in a subset is at least 4 or 5 (for locating in three dimensions) and they are not coplanar (Schmidt, 1972).

The localizer that has been implemented here employs a gradient-descent procedure and solves an unconstrained quadratic minimization problem to determine the intersection point of a locus of hyperboloids satisfying a given range difference. The exact procedure is not important, and any of a number of methods found in the literature can be employed. Nevertheless it should be kept in mind that as the number of sources and sensors becomes large, algorithm efficiency and computational speed become important. It may become necessary to refine the procedure by improving the initial estimate to bring it close to the global minimum (as suggested in Smith and Abel, 1987) or by pre-processing the data to select only those microphones that are closest to the source (i.e., microphones where the signal has the most energy). Further, once the set of single-source data frames have been identified, the localization can be parallelized to simultaneously extract all source locations as these operations are independent of one another. This report has not exmained issues of convergence nor has it tried alternate procedures or computational schemes. These are important problems for future work.

Certain factors constrain localization accuracy and the ability to resolve two sources located close to one another. For a single source, localization accuracy reduces as the source-to-array distance increases. Failure to accurately resolve two sources that are close to one another also occurs when the distance from the sources to the array increases. This is due to a reduction in the angular separation and hence relative time delay between two sources (as viewed from one of the sensors) with increasing distance, leading to misiden-

tifying both callers as belonging to one source. Increasing the array aperture by spreading out the microphones can improve this, but the increased relative delay may reduce the beamformer's ability to reject other sources during recovery, as is discussed further below. Enlarging the array by adding more sensors over more area can overcome both problems and can be used to cover arbitrarily large choruses. These tests are on-going.

The effect of source-sensor distance and angular separation between sources can be seen in the reported results. In Fig. 6(A), cricket frogs 5 and 6 are much further away from the sensors than the remaining frogs. The location estimates for these frogs are more variable than for the remaining clusters (standard deviations are reported in Table I). Similarly the toad cluster 4 shown in Fig. 6(B) represents a cluster of two toads. The area of analysis was a square 9 m wide. The cluster is shown at the right boundary ($x=9$) but this is because the algorithm projects all sources outside the selected area of analysis on to the boundary. In actuality the x-coordinate exceeded 9 m (see Table I). Increasing the area of analysis beyond the 9 m^2 did not improve the accuracy nor did it resolve the two toads into their component sources (analysis not shown). In this case the toads could have been resolved by adding more microphones to the right of mics 2 and 3.

These examples illustrate that array geometry is not critical to the analyis, but it is important to provide adequate spatial coverage of the section of the chorus of interest. In this study, four microphones covered about 100 sq m, but this number depends on the elevation of the microphones and required spatial resolution. The preferred array geometry, placement, and trade-offs warrant further investigation, but the proposed methods can be applied to any microphone configuration.

The clustering procedure assigns a set of estimated locations to a given caller (i.e., the data frames $P_i \in P$ to a single source i). The procedure is valid if the variance of source-location estimates are small in comparison with the inter-source spacing, and if the source is spatially stationary on the time-scale over which the cluster is determined. The data on the within-location and across-location variability can be inferred from Table I. The callers did not exhibit significant movement over a cluster interval of about 1 min, and so this was the interval employed in the study. In other situations or for other species, the clustering interval may need to be established by trial and error before selecting a suitable time-frame for analysis.

Clusters were evaluated visually. This is readily performed even for several hours of data because it involves the selection of a bounding box for each individual, as shown in Fig. 5. However, an automated clustering algorithm can be implemented, for instance, by examining the histogram of locations for peaks, although it has not been tried here. This is an area for future work.

C. Estimation of steering vectors

To recover the individual vocalizations with little attenuation or distortion, the steering vectors must be estimated accurately. This was accomplished by a novel blind field steering-vector estimation that utilized only those time-frames where a single source was present. These frames were obtained from the localization step. The clustering procedure then assigned a unique source to every cluster. The steering vectors were evaluated from the sensor covariance matrix averaged over the cluster, and then they were normalized with respect to the microphone where the source had greatest power. The method provides a fast and accurate estimate of the steering vector.

D. Adaptive beamforming and source recovery

With known steering vectors the adaptive beamformer output recovers the individual frog vocalizations with little attenuation or distortion. While steering vectors are estimated only in the selected frames of data, the beamformer filters the entire data set based on the assumption that the steering vector does not change between frames. In other words, it makes the biologically plausible assumption that the source does not move in the intervening time intervals.

For widely separated frogs, there is little energy from the other callers in the beamformer output. However, nearby frogs are often only partially attenuated, resulting in interference (cross-talk) within a recovered channel. While the extent of cross-talk resulting from interfering sources is a measure of the beamformer performance, it is not easily evaluated because of the adaptive nature of the spatal filter (see Haykin, 2002, for a discussion). The performance depends on a number of factors including the steering vector, the temporal and spectral characteristics of the target and interfering sources, and the array geometry. Performance is therefore highly dependent on the context.

To illustrate the beamformer capabilities and some of the factors that can influence its performance, simulations were carried out with two sources located over a range of distances (Fig. 7). The recovery of the target demonstrates little or no self-cancellation [Fig. 7(A), ○] and the extent of cross-talk is small (<-20 dB) for source separations larger than about 3–4 cm [Fig. 7(A), □]. While the beamwidth is sharp, it reaches an apparent minima at about 0.1 m. This minima is due to the tonal nature of the *Bv* call which has a spectral peak around 1700 Hz corresponding to a half wavelength $\lambda/2 \approx 0.1$ m. At this separation the distinction between the steering vectors of the target and interferer is maximum and results in maximal attenuation or minimal cross-talk. For greater source separation, the attenuation reduces progressively because the increasing relative delays between microphones increase the intrinsic time-domain length of the optimal spatial cancellation filters. For practical reasons, these filters are limited to a (fixed) FFT length, and therefore there is an effective truncation of the filters that limits the beamformer performance.

Research into biologically inspired binaural beamformers for hearing aids has led to new methods that exploit some of the mechanisms found in the auditory system (Liu *et al.*, 2000; Lockwood *et al.*, 2004). These mechanisms enhance the performance of the beamfomer even with two micro-

phones in complex cocktail-party scenes (Cherry, 1953) containing many simultaneous speech sources. This is a situation similar to that of a frog chorus.

A chorus contains a high density of spatially localized nonstationary interfering sources that exhibit time-frequency sparseness. That is, at any given instant energy is present only in a small region of time-frequency space. By exploiting the time-frequency sparseness of the target and interfering sources, the complex scene could be separated into individual time-frequency bins with much less overlap (Mohan *et al.*, 2008). This makes it easier to identify the direction of sources and to beamform independently in these sparse channels to obtain improved cancellation of interference. The most efficient of these beamformers for small arrays, a particular frequency-domain MVDR beamformer implementation, combines very rapid independent adaptation in each time-frequency bin with low computational complexity (Lockwood *et al.*, 2004). This implementation is particularly suited to the complex dynamics of a frog chorus. The rapidly varying nonstationarity and time-frequency sparseness of the frog vocalizations, the presence of many more frogs than microphones, and their small spatial separation combine to make the use of this particular adaptive beamformer very appropriate.

E. Monitoring choruses and future directions

The procedures outlined in this work can be used to monitor heterospecific choruses where the number of individuals exceeds the number of microphones. A total of 9 individuals from two species were localized. Large natural choruses can involve many more frogs and several species in breeding sites that exceed several thousand sq m. While there is variability in the size of natural choruses reported (see Sec. I A), there is no doubt that the spatial extent of the chorus and the number of individuals reported here are small compared to many natural choruses. Thus, it is of interest to ask whether the methods can be applied to large choruses, and what are the limits on the chorus sizes that can be analyzed.

There are no clear answers to these questions at this time although they form the focus of on-going work. However, several features of the processing scheme should be noted in this regard.

(1) The array is scalable. That is, more microphones can be added to increase the spatial extent of the array, and increase coverage of larger choruses. For example, in the scheme presented here, two additional microphones (say, mics 5 and 6) could have been added to the right of mics 3 and 4 to increase the coverage beyond $x=9$ m. The size of the array is only limited by practical considerations. The methods presented here are independent of the array size or geometry.
(2) The analysis of a chorus does not require data from all microphones. As discussed earlier, there is much greater advantage in focusing or applying a spotlight to a neighborhood around a set of microphones. At a practical level this would mean ignoring data from microphones that are much further away from the neighborhood. A group of five microphones deployed as one module (in the configuration of an irregular polyhedron) would be adequate to cover a neighborhood around the module and localize in three-dimensions. More modules could be added to increase the array size.
(3) The number of callers localized and separated (9) exceeded the number of microphones (4). This is a major advance, as blind source separation requires as many sensors as there are sources (Hyvarinen *et al.*, 2001). By exploiting the time-frequency sparseness of the system and applying biologically motivated strategies, the processor is able to separate more sources than sensors. We have no data as yet on the upper limit on the number of sources that can be extracted by a fixed number of microphones. These tests are on-going. If we were to increase the number of microphone modules as suggested above, then in principle the processor could analyze larger choruses. This work is a step in that direction.

A significant drawback of the method is that it does not include multi-path propagation. As a result the cross-correlation function between pairs of sensors will demonstrate multiple peaks, and it may not be possible to determine the correct time delay without errors. We note that only time-delay estimation via cross-correlation is affected; the localizer and the beamformer are unaffected by multi-path propagation. Thus, a major future goal is to estimate the arrival-time differences to the microphones by incorporating multi-path propagation.

ACKNOWLEDGMENTS

The authors thank Margaret Jones (University of Chicago) and Kimberly Walsh (University of Texas at San Antonio) for assistance in data collection and data analysis. They also thank Caroline Chipman Evans and Jan Wrede at the Cibolo Nature Center for use of the site to collect data, and for their enthusiastic support of the project. The research was supported by Grant No. CCR 03-12432 ITR from the National Science Foundation (to D.L.J.), Grant No. R01 DC005762-02 from the National Institute on Deafness and Other Communication Disorders (to D.L.J.), the Norman Hackerman Advanced Research Program (NHARP) Grant No. 010115-0101-2007 from the Texas Higher Education Coordinating Board (to R.R.), and UTSA STARS funds (to R.R.). They thank the anonymous reviewer and the handling editor Dr. Michael Owren for their comments and criticisms on the manuscript.

[1]The minimum number of sensors that is required was derived by Schmidt (1972) in his "Location on the conic axis" or LOCA method. Briefly, and in two-dimensions, three sensors are assumed to be located on a generalized conic with the source located at one of the foci. The eccentricity of the solution conic determines whether the conic is an ellipse, a hyperbola, or a parabola. If the conic is an ellipse, then the three sensors will unambiguously locate the source at one of the foci with the other foci yielding the negative of the time-difference measurements. In case the conic is a hyperbola, the foci cannot be diambiguated because they generate the same time-difference measurement. In this case a fourth sensor is necessary to uniquely locate the source. In the limiting case of a parabola, one of the foci will be at infinity. The extension to three dimensions is similar. Hence, the minimum number of sensors that are required will depend on the geometry of the source-sensor arrangement. The LOCA method should

be contrasted with the hyperbolic range difference location method of van Etten (1970) where the sensors are at the foci and the source is at the intersection of the hyperboloids. The two methods are mathematical duals (Schmidt, 1972).

[2]The use of a global positioning system for monitoring sensor positions or for acoustic localization has not been reviewed here, although this is a technology that is likely to see widespread use in the future.

Banks, D. (**1993**). "Localization and separation of simultaneous voices with two microphones," IEE Proc.-I Commun. Speech Vision **140**, 229–234.

Bee, M. A., and Micheyl, C. (**2008**). "The cocktail party problem: What is it? How can it be solved? And why should animal behaviorists study it?" J. Comp. Psychol. **122**, 235–251.

Blair, W. F. (**1958**). "Mating call in the speciation of anuran amphibians," Am. Nat. **92**, 27–51.

Blauert, J. (**1983**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

Bodden, M. (**1993**). "Modeling human sound source localization and the cocktail party effect," Acta Acust. (Beijing) **1**, 43–55.

Bogert, C. M. (**1960**). "The influence of sound on the behavior of amphibians and reptiles," in *Animal Sounds and Communication*, edited by W. E. Lanyon and W. N. Tavolga (American Institute of Biological Sciences, Washington, DC), pp. 137–320.

Bower, J. L., and Clark, C. W. (**2005**). "A field test of the accuracy of a passive acoustic location system," Bioacoustics **15**, 1–14.

Bradbury, J. W. (**1981**). "The evolution of leks," in *Natural Selection and Social Behavior: Recent Research and New Theory*, edited by R. D. Alexander and D. W. Tinkle (Chiron, New York), pp. 138–169.

Brandstein, M., and Ward, D. (**2001**). *Microphone Arrays: Signal Processing Techniques and Applications* (Springer, New York).

Brush, J. S., and Narins, P. M. (**1989**). "Chorus dynamics of a neotropical amphibian assemblage: Comparison of computer simulation and natural behavior," Anim. Behav. **37**, 33–44.

Capon, J. (**1969**). "High-resolution frequency-wavenumber spectrum analysis," Proc. IEEE **57**, 1408–1419.

Capranica, R. R. (**1965**). in *The Evoked Response of the Bullfrog: A Study of Communication in Anurans*, Research Monograph No. 33 (MIT, Cambridge, MA).

Carter, G. C. (**1981**). "Guest editorial: Time delay estimation," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-29**(3), 461–462.

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and two ears," J. Acoust. Soc. Am. **25**, 975–979.

Clark, C. W. (**1980**). "A real-time direction finding device for determining the bearing to the underwater sounds of southern right whales *Eubalaena australis*," J. Acoust. Soc. Am. **68**, 508–511.

Clark, C. W., (**1989**). "Call tracks of bowhead whales based on call characteristics as an independent means of determining tracking parameters," Rep. Int. Whal. Comm. **39**, 111–113.

Clark, C. W., Ellison, W. T., and Beeman, K. (**1986**). "Acoustic tracking of migrating bowhead whales," Proc. IEEE Oceans '86 (IEEE, New York), pp. 341–346.

Clark, C. W., Charif, R., Mitchell, S., and Colby, J. (**1996**). "Distribution and behavior of the bowhead whale, *Balaena mysticetus*, based on analysis of acoustic data collected during the 1993 spring migration off Point Barrow, Alaska," Rep. Int. Whal. Comm. **46**, 541–552.

Cox, H., Zeskind, R. M., and Kooij, T. (**1986**). "Practical supergain," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-34**, 393–398.

Cox, H., Zeskind, R. M., and Owen, M. M. (**1987**). "Robust adaptive beamforming," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**, 1365–1376.

Delosme, J. M., Morf, M., and Friedlander, B. (**1987**). "Source location from time differences of arrival: Identifiability and estimation," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**, 818–824.

Deslodge, J. G. (**1998**). "The location-estimating null-steering (LENS) algorithm for adaptive microphone array-processing," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Ehret, G., and Gerhardt, H. C. (**1980**). "Auditory masking and the effects of noise on the responses of the green treefrog (*Hyla cinerea*) to synthetic mating calls," J. Comp. Physiol. [A] **141**, 1–12.

Feng, A. S., and Ratnam, R. (**2000**). "Neural basis of hearing in real-world situations," Annu. Rev. Psychol. **51**, 699–725.

Friedl, T. W. P., and Klump, G. M. (**2005**). "Sexual selection in the lek-breeding European treefrog: Body size, chorus attendance, random mating and good genes," Anim. Behav. **70**, 1141–1154.

Frost, O. L. (**1972**). "An algorithm for linearly constrained adaptive array processing," Proc. IEEE **60**, 926–935.

Gaik, W. (**1993**). "Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling," J. Acoust. Soc. Am. **94**, 98–110.

Gerhardt, H. C. (**1994**). "The evolution of vocalizations in frogs and toads," Annu. Rev. Ecol. Syst. **25**, 293–324.

Gerhardt, H. C., and Huber, F. (**2002**). *Acoustic Communication in Insects and Anurans* (University of Chicago Press, Chicago, IL).

Gerhardt, H. C., and Klump, G. M. (**1988**). "Masking of acoustic signals by the chorus background noise in the green treefrog: A limitation on mate choice," Anim. Behav. **36**, 1247–1249.

Gerhardt, H. C., Klump, G. M., Diekamp, B., and Ptacek, M. (**1989**). "Intermale spacing in choruses of the spring peeper, *Pseudacris (Hyla) crucifer*," Anim. Behav. **38**, 1012–1024.

Grafe, T. U. (**1997**). "Costs and benefits of mate choice in the lek-breeding reed frog, *Hyperolius marmoratus*," Anim. Behav. **53**, 1103–1117.

Griffiths, L. J., and Jim, C. W. (**1982**). "An alternative approach to linearly constrained adaptive beamforming," IEEE Trans. Antennas Propag. **AP-30**, 27–34.

Hahn, W. R., and Tretter, S. A. (**1973**). "Optimum processing for delay-vector estimation in passive arrays," IEEE Trans. Inf. Theory **IT-12**, 608–614.

Haykin, S. (**2002**). *Adaptive Filter Theory*, 4th ed. (Prentice-Hall, Englewood Cliffs, NJ).

Hyvarinen, A., Karhunen, J., and Oja, E. (**2001**). *Independent Component Ananlysis* (Wiley, New York).

Jeffress, L. A. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**, 35–39.

Konishi, M. (**1992**). "The neural algorithm for sound localization in the owl," Harvey Lect. **86**, 47–64.

Kroodsma, D. E., Miller, E. H., and Ouellet, H. (**1983**). *Acoustic Communication in Birds, Vol. 2: Song Learning and Its Consequences* (Elsevier Science and Technology, New York).

Lindemann, W. (**1986**). "Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals," J. Acoust. Soc. Am. **80**, 1608–1622.

Liu, C., Wheeler, B. C., O'Brien Jr., W. D., Bilger, R. C., Lansing, C. R., and Feng, A. S. (**2000**). "Localization of multiple sound sources using two microphones," J. Acoust. Soc. Am. **108**, 1888–1905.

Lockwood, M. E., and Jones, D. L. (**2006**). "Beamformer performance with acoustic vector sensors in air," J. Acoust. Soc. Am. **119**, 608–619.

Lockwood, M. E., Jones, D. L., Bilger, R. C., Lansing, C. R., O'Brien Jr., W. D., Wheeler, B. C., and Feng, A. S. (**2004**). "Performance of time- and frequency-domain binaural beamformers based on recorded signals from real rooms," J. Acoust. Soc. Am. **115**, 379–391.

Magyar, I., Schleidt, W. M., and Miller, D. (**1978**). "Localization of sound producing animals using the arrival time differences of their signals at an array of microphones," Experientia **34**, 676–677.

McGregor, P. K., Dabelsteen, T., Clark, C. W., Bower, J. L., Tavares, J. P., and Holland, J. (**1997**). "Accuracy of a passive acoustic location system: Empirical studies in terrestrial habitats," Ethol. Ecol. Evol. **9**, 269–286.

Mennill, D. J., Burt, J. M., Fristrup, K. M., and Vehrencamp, S. L. (**2006**). "Accuracy of an acoustic location system for monitoring the position of duetting songbirds in tropical forest," J. Acoust. Soc. Am. **119**, 2832–2839.

Mohan, S., Lockwood, M. E., Kramer, M. L., and Jones, D. L. (**2008**). "Localization of multiple acoustic sources with small arrays using a coherence test," J. Acoust. Soc. Am. **123**, 2136–2147.

Murphy, C. G. (**1994**). "Chorus tenure of male barking tree frogs, *Hyla gratiosa*," Anim. Behav. **48**, 763–777.

Murphy, C. G. (**2003**). "The cause of correlations between nightly numbers of male and female barking treefrogs (*Hyla gratiosa*) attending choruses," Behav. Ecol. **14**, 274–281.

Narins, P. M. (**1982**). "Effects of masking noise on evoked calling in the Puerto Rican coqui frog (Anura: Leptodactylidae)," J. Comp. Physiol. [A] **147**, 439–446.

Perrill, S. A., and Shepherd, W. J. (**1989**). "Spatial distribution and male-male communication in the Northern Cricket Frog, *Acris crepitans blanchardi*," J. Herpetol. **23**, 237–243.

Schau, H. C., and Robinson, A. Z. (**1987**). "Passive source localization employing intersecting spherical surfaces from time-of-arrival differences," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**, 1223–1225.

Schmidt, R. O. (**1972**). "A new approach to geometry of range difference location," IEEE Trans. Aerosp. Electron. Syst. **AES-8**, 821–835.
Schwartz, J. J. (**2001**). "Call monitoring and interactive playback systems in the study of acoustic interactions among male anurans," in *Anuran Communication*, edited by M. J. Ryan (Smithsonian, Washington, DC), pp. 183–204.
Schwartz, J. J., and Gerhardt, H. C. (**1995**). "Directionality of the auditory system and call pattern recognition during acoustic interference in the gray treefrog *Hyla versicolor*," Am. Nat. **1**, 195–206.
Schwartz, J. J., and Wells, K. D. (**1983a**). "An experimental study of acoustic interference between two species of neotropical treefrogs," Anim. Behav. **31**, 181–190.
Schwartz, J. J., and Wells, K. D. (**1983b**). "The influence of background noise on the behavior of a neotropical treefrog *Hyla ebraccata*," Herpetologica **39**, 121–129.
Simmons, A. M., Popper, A. N., and Fay, R. R. (**2002**). *Acoustic Communication*, Springer Handbook of Auditory Research (Springer, New York).
Simmons, A. M., Simmons, J. A., and Deligeorges, S. A. (**2006**). "Temporal organization of bullfrog choruses," J. Acoust. Soc. Am. **119**, 3210.
Simmons, A. M., Simmons, J. A., and Bates, M. E. (**2008**). "Analyzing acoustic interactions in natural bullfrog choruses," J. Comp. Psychol. **122**, 274–282.
Smith, J. O., and Abel, J. S. (**1987**). "Closed-form least-squares source location estimation from range-difference measurements," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**, 1661–1669.
Spiesberger, J. L. (**1998**). "Linking auto- and cross-correlation functions with correlation equations: Application to estimating the relative travel times and amplitudes of multipath," J. Acoust. Soc. Am. **104**, 300–312.
Spiesberger, J. L. (**1999**). "Locating animals from their sounds and tomography of the atmosphere: Experimental demonstration," J. Acoust. Soc. Am. **106**, 837–846.
Spiesberger, J. L., and Fristrup, K. M. (**1990**). "Passive localization of calling animals and sensing of their acoustic environment using acoustic tomography," Am. Nat. **135**, 107–153.
Stewart, M. M., and Pough, F. H. (**1983**). "Population density of tropical forest frogs: Relation to retreat sites," Science **221**, 570–572.
Sullivan, B. K., and Wagner, W. E., Jr. (**1988**). "Variation in advertisement and release calls, and social influences on calling behavior in the Gulf Coast Toad (*Bufo valliceps*)," Copeia **1988**, 1014–1020.
Sullivan, B. K., Ryan, M. J., and Verrell, P. A. (**1995**). "Female choice and mating system structure," in *Amphibian Biology, Vol. 2: Social Behaviour*, edited by H. Heatwole and B. K. Sullivan (Chipping Norton, Surrey Beatty), pp. 469–517.
van Etten, J. P. (**1970**). "Navigation systems: Fundamentals of low and very-low frequency hyperbolic techniques," Electrochem. Commun. **45**, 192–212.
van Veen, B. D., and Buckley, K. M. (**1988**). "Beamforming: A versatile approach to spatial filtering," IEEE ASSP Mag. **5**, 4–24.
Wagner, W. E., Jr., and Sullivan, B. K. (**1992**). "Chorus organization in the Gulf Coast Toad (*Bufo valliceps*): Male and female behavior and the opportunity for sexual selection," Copeia **1992**, 647–658.
Watkins, W. A., and Schevill, W. E. (**1972**). "Sound source locations by arrival times on a non-rigid three-dimensional hydrophone array," Deep-Sea Res. **19**, 691–706.
Wax, M., and Kailath, T. (**1983**). "Optimum localization of multiple sources by passive arrays," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-31**, 1210–1217.
Wells, K. D. (**1977**). "The social behavior of anuran amphibians," Anim. Behav. **25**, 666–693.
Wilczynski, W., and Brenowitz, E. A. (**1988**). "Acoustic cues mediate intermale spacing in a neotropical frog," Anim. Behav. **36**, 1054–1063.
Wollerman, L. (**1999**). "Acoustic interference limits call detection in a neotropical frog Hyla ebraccata," Anim. Behav. **57**, 529–536.
Yin, T. C. T., and Chan, J. C. K. (**1990**). "Interaural time sensitivity in medial superior olive of cat," J. Neurophysiol. **64**, 465–488.
Zelick, R., and Narins, P. M. (**1985**). "Characterization of the advertisement call oscillator in the frog, *Eleutherodactylus coqui*," J. Comp. Physiol. [A] **156**, 223–229.

The effect of altered auditory feedback on control of vocal production in budgerigars (*Melopsittacus undulatus*)

Michael S. Osmanski[a] and Robert J. Dooling
Department of Psychology, University of Maryland, College Park, Maryland 20742

(Received 18 August 2008; revised 29 May 2009; accepted 29 May 2009)

Budgerigars learn their vocalizations by reference to auditory information and they retain the ability to learn new vocalizations throughout life. Auditory feedback of these vocalizations was manipulated in three experiments by training birds to produce vocalizations while wearing small earphones. Experiments 1 and 2 examined the effect of background noise level (Lombard effect) and the effect of manipulating feedback level from self-produced vocalizations (Fletcher effect), respectively. Results show that birds exhibit both a Lombard effect and a Fletcher effect. Further analysis showed that changes in vocal intensity were accompanied by changes in call fundamental frequency and duration. Experiment 3 tested the effect of delaying or altering auditory feedback during vocal production. Results showed subsequent production of incomplete and distorted calls in both feedback conditions. These distortions included changes in the peak fundamental frequency, amplitude, duration, and spectrotemporal structure of calls. Delayed auditory feedback was most disruptive to subsequent calls when the delay was 25 ms. Longer delays resulted in fewer errors.
 [DOI: 10.1121/1.3158928]

PACS number(s): 43.80.Ka, 43.80.Lb [ADP] Pages: 911–919

I. INTRODUCTION

Animals that learn their vocalizations rely on auditory feedback (AF) for the development and maintenance of a normal vocal repertoire (for reviews, see Farabaugh and Dooling, 1996; Janik and Slater, 1997; Doupe and Kuhl, 1999; Brainard and Doupe, 2000; Janik and Slater, 2000; and Boughman and Moss, 2003). AF during vocal production has been most extensively studied in humans, where feedback mechanisms help regulate, among other things, vocal amplitude. The Lombard effect, for instance, describes an increase in vocal amplitude in response to an increase in ambient noise level (Lombard, 1911; for a review, see Lane and Tranel, 1971). The Fletcher effect describes a decrease in vocal amplitude in response to an increase in perceived vocal loudness (Fletcher *et al.*, 1918; Lane and Tranel, 1971; Siegel and Pick, 1974). In human speakers, these changes also include increases in syllable duration and vocal pitch and decreases in speaking rate (Hanley and Steer, 1949; Draegert, 1951; Dreher and O'Neill, 1958). Presumably, these responses function to preserve speech intelligibility in varied listening conditions.

Signal degradation due to environmental noise is a problem for all acoustic communication systems. Mechanisms of noise-dependent amplitude changes in vocal behavior similar to the Lombard effect have been described in monkeys (Sinnott *et al.*, 1975; Brumm *et al.*, 2004; Egnor and Hauser, 2006), quails (Potash, 1972), hummingbirds (Pytte *et al.*, 2003), songbirds (Cynx *et al.*, 1998; Brumm and Todt, 2002; Kobayashi and Okanoya, 2003), and budgerigars (Manabe *et al.*, 1998). Generally, in animal studies, technical challenges in delivering noise or altered AF make precise comparisons with human work difficult. The present studies use small earphones to overcome this limitation.

Changes in vocal output caused by altered feedback extend beyond level effects. Precise timing of feedback also has important consequences for the normal development and maintenance of vocal production. In humans, delayed auditory feedback (DAF) of the speech signal results in a number of disruptive effects, including slower speech rate, higher fundamental frequency, longer syllable durations, and a range of production errors (including stuttering and short consonant-like bursts of sound) while some subjects report a complete inability to continue speaking (Lee, 1950; Fairbanks, 1955; Yates, 1963; Howell and Archer, 1984). The most severe disruptions in speech occur at a feedback delay of about 200 ms, with less disruption at shorter and longer delays. Interestingly, DAF has fluency-enhancing effects on stutterers (e.g., Bloodstein, 1995). Taken together, this suite of effects has been interpreted as evidence for timing malfunctions in a closed-loop feedback circuit, which controls ongoing vocal production via AF (e.g., Fairbanks, 1954; Chase, 1965).

There is very little DAF work in animals. Recent work in songbirds has examined the effects of manipulating AF by playing either altered or delayed song during vocal production in zebra finches (Leonardo and Konishi, 1999; Cynx and Von Rad, 2001; Sakata and Brainard, 2006) resulting in dramatic effects on song. Birds show song syllable repetition, syllable deletion, and loss of syllable sequencing and structure under these conditions. The most severe disruptions caused by DAF occurred at delays of 100 ms in zebra finches (Cynx and Von Rad, 2001) and about 65 ms in Bengalese finches (Sakata and Brainard, 2006).

[a]Author to whom correspondence should be addressed. Present address: Department of Biomedical Engineering, Johns Hopkins University School of Medicine, 412 Traylor Research Building, 720 Rutland Avenue, Baltimore, MD 21205. Electronic mail: michael.osmanski@jhu.edu

Here the authors apply new methods to a nonsongbird species to further examine how altered AF affects vocal production in birds and also to provide insights into the operation of auditory-vocal circuits in budgerigars. First, vocal behavior was rigorously controlled by training birds, through operant conditioning with food reward, to produce specific vocalizations (i.e., contact calls) to a visual cue. Second, birds were tested while wearing small earphones allowing more precise delivery of noise or altered AF. In experiment 1 (Lombard effect), the authors played various levels of white noise to subjects through the earphones while they were vocalizing. In experiment 2 (Fletcher effect), the authors played amplitude-adjusted exemplars of the birds' contact calls through the earphones as the birds were producing the same vocalizations. In experiment 3, the authors examined the effect of spectrotemporal alterations of AF (e.g., DAF, reversing the bird's call) on vocal production.

II. GENERAL METHODS

A. Subjects

The subjects in all three experiments were three adult male budgerigars from a colony maintained in an aviary at the University of Maryland. Each bird was separately caged and had *ad libitum* access to water. Since food was used to reinforce vocal behavior, the birds were maintained at 90% of their free-feeding body weight. The University of Maryland Animal Care and Use Committee approved all experimental procedures.

B. Apparatus

Birds were trained in an operant testing apparatus consisting of a small wire cage ($14 \times 12 \times 17$ cm^3) constructed of wire mesh and mounted in an acoustic isolation chamber (Industrial Acoustic Co. model AC-1). Three light-emitting diodes (LEDs) (left, center, and right) were attached to a piece of anechoic foam on the front panel of the cage at approximately the level of the birds' heads. Three small speakers (SONY model MDR-Q22LP) were mounted on the exterior of the cage—one at the center above the front LED panel and one on each of the left and right sides. A small directional microphone (SONY model ECM-77B) located just below the LED panel detected vocalizations. A food hopper containing hulled millet was located on the floor of the cage under the front LED panel. A small video camera was used to monitor the bird's behavior while in the chamber.

C. Training/testing procedure and analysis

1. Contact call detection and analysis

Training, testing, and analysis programs were written in MATLAB software (version 6.5, Natick, MA) for Tucker Davis Technologies (TDT) System III hardware (Gainesville, FL). The output of the microphone was amplified, low-pass filtered at 10 kHz, and sent to a circular memory buffer in a TDT real-time digital signal processor (RP2.1) at a sampling rate of 25 kHz. A typical budgerigar contact call duration is 100–150 ms with spectral energy concentrated between 2 and 4 kHz (Farabaugh *et al.*, 1994; Farabaugh and Dooling, 1996; Farabaugh *et al.*, 1998). Thus, incoming signals were classified as contact calls if signal intensity exceeded a user-defined value for a minimum of 70 ms and signal power in the frequency band between 2 and 4 kHz exceeded that between 4 and 10 kHz.

All signals classified as contact calls were saved for later analysis. Analysis first involved the generation of serial power spectra across each call in 5 ms (i.e., 122 pt) windows (with 50% window overlap) using a chirp-z transform spectral estimation method (MATLAB function CZT), which allowed 1 Hz frequency resolution. Within each window, the authors measured the frequency and amplitude of the spectral peak and also the spectral bandwidth 3 dB down from the peak. These measures were then averaged to derive the average peak frequency, peak amplitude, and 3 dB bandwidth across the call. Finally, the authors measured the call duration and calculated the similarity to the stored standard or template call (see Secs. 3 and 4 for a description of the template and correlation algorithm). These measures were later analyzed using SPSS software (version 12.0, Chicago, IL).

2. Initial training (shaping)

Birds were habituated to the experimental chamber and trained to eat from the food hopper when it was activated. Once the birds consistently ate from the raised hopper, shaping of vocal production began. A tape recording of a flock of vocalizing budgerigars was played in the operant chamber to induce the birds to vocalize. Whenever the birds responded to this flock tape with a contact call, the experimenter activated the hopper. Birds quickly came to associate vocalizing in the test chamber with delivery of food. The flock tape playback became unnecessary after several training sessions as the bird reliably produced calls to obtain food and food reinforcement was delivered automatically under computer control.

Birds were next trained to vocalize only when the center LED was illuminated. The LED was turned off once a vocalization was acquired and then turned on again after a random time interval (approximately 5–15 s). Only vocalizations produced when the light was illuminated were reinforced. Vocalizations produced when the LED was turned off caused the random interval timer to reset and thus increased the wait time before another trial could be started (i.e., the LED turned back on). Birds quickly learned to vocalize in the chamber only in response to the illumination of the center LED.

3. Selecting the contact call template

Once the birds were reliably responding, they were run in several additional training sessions to establish their call repertoire. Budgerigars produce several different call-types that are easily distinguished based on spectrographic characteristics, with one call-type typically being produced much more often than others (Farabaugh *et al.*, 1994; Farabaugh

and Dooling, 1996). An exemplar of the bird's most frequent contact call-type was selected as that bird's standard or template call (see Manabe and Dooling, 1997).

The template call was chosen by first computing pairwise spectral cross-correlations among all calls in the test sessions. A custom MATLAB program created a spectrogram for each call using a 256-point (i.e., 10.5 ms) Hanning window with 50% window overlap. Spectrograms were compared using two-dimensional cross-correlation (MATLAB function XCORR2), resulting in a series of correlation values representing all possible temporal offsets between the two spectrograms. The maximum correlation value was taken as the similarity index between the two calls and was normalized to $r=0.0$ if the two calls were perfectly dissimilar and $r=1.0$ if the calls were identical. The resultant similarity matrix was analyzed using a multidimensional scaling algorithm (MATLAB function MDSCALE) and the call in the center of the largest cluster in this two-dimensional space was selected as the template call for the next phase of training.

4. Training vocal precision

Subsequent training sessions used the template call described above to differentially reinforce the bird's vocal behavior by only reinforcing calls that were similar to the template call. Every vocalization produced by a bird was compared to the stored digital template in real-time and the bird was reinforced if the correlation between the two calls exceeded an experimenter-defined criterion. At first, the criterion correlation value was set very low (e.g., $r=0.01$) so that all calls were reinforced. The criterion was gradually increased over several sessions to a maximum value of $r=0.70$. All training sessions were terminated after 50 reinforcements or 25 min, whichever came first. Birds almost always completed 50 reinforcements within 25 min (>95% of sessions). Subjects were tested in two daily sessions, 5 days/week. All test sessions were separated by at least 3 h.

D. Earphones

1. Surgical procedure and earphone construction

Once the birds were trained, a small, stainless steel head post (jewelry pin with clutch back, Hirschberg Schutz & Co., Inc., model no. JC8425–01) was affixed to each bird's skull. First, the animal was weighed and given an intramuscular injection of ketamine (40 mg/kg)/xylazine (20 mg/kg). The toe pinch response was used to determine whether the bird was properly anesthetized for surgery. Next, the superior aspect of the skull was exposed using a No. 11 scalpel blade and Vanass scissors (Fine Science Tools, Foster City, CA). The skull surface was abraded using the scalpel to create better adhesion before the head post was attached using dental cement (A-M Systems Inc.). Nexaband (Closure Medical Corporation, Raleigh, NC) was used to seal the incision, and the bird was placed in a heated therapy unit for monitoring until the anesthetic effects had worn off. Birds were monitored for 24–48 h following surgery and a non-narcotic, nonsteroidal analgesic (Flunixin meglumine, 10 mg/kg) was administered daily during this recovery period.

Following recovery, birds were fitted with an earphone assembly. The earphone frame was constructed using thin steel wire (1 mm diameter) with small rubber grommets (10 mm diameter) as earphone cushions. A transducer (Knowles Acoustics, model no. EH-3062) was glued to the interior of each grommet using commercially available silicone sealant. These transducers have a frequency response from 0.2 to 8 kHz with peak sensitivities between 2 and 5 kHz. When affixed to the head post, the transducers were aligned directly at the opening of the bird's ear canal. The grommets pressed lightly against the sides of the bird's head providing some attenuation of external sounds. During testing, wires from the earphones were fed through the ceiling of the operant chamber to the output amplifiers of the TDT hardware system. The birds were able to move around freely while wearing the headphones during a test session. The headphones were attached to the head post prior to testing and removed after the test session was complete when the bird was returned to the aviary.

2. Earphone calibration and testing

The sound pressure level of the feedback was measured with a Larson-Davis model 824 sound level meter and 3-m extension cable with a $\frac{1}{4}$ in. microphone both before and after the experiment. The microphone was placed inside a custom-made open adaptor, which approximated the diameter of the bird's auditory meatus and the distance to the bird's tympanic membrane from the transducer.

After being fitted with earphones, each bird was tested in several training sessions to ensure that performance was not affected by the surgery, wearing earphones, or the presence of wiring above the bird's head. No sounds were delivered through the earphones during these sessions and all birds achieved and maintained a reinforcement rate greater than 90% within five sessions after being reintroduced into the testing environment following surgery. The possibility of occlusion effects was considered unlikely because the earphones were not tightly pressed against the head, the energy in the contact calls was at relatively high frequencies between 2 and 4 kHz, and budgerigars have an interaural pathway that connects their middle ears.

III. EXPERIMENT 1: LOMBARD EFFECT

The Lombard effect is well studied in humans and has been shown in a number of nonhuman animals, including budgerigars. Here the authors examined the Lombard effect in budgerigars with noise delivered through earphones over a broad range of noise levels. In keeping with free field work (Manabe *et al.*, 1998) the authors hypothesized that vocal amplitude would increase as the level of the AF increased. The authors also hypothesized, based on work with human speech, that increasing vocal amplitude would be accompanied by parallel increases in fundamental frequency and duration.

A. Methods

1. Subjects

Three adult male budgerigars were used in this experiment.

B. Procedure

Once the birds were trained to asymptotic levels of performance on the template-training task described above and were fitted with earphones, Gaussian white noise from the TDT System III RP2.1 hardware was delivered during testing. The feedback noise level was measured using a Larson-Davis model 824 sound level meter for 11 different noise levels [40–90 dB sound pressure level (SPL) in 5 dB steps, *A*-weighting, fast rms] at the bird's ear.

Birds were tested in four sessions of 60 trials each. A trial was defined by a bird producing a single vocalization in response to an illuminated LED. All 11 noise levels (and a quiet condition) were presented across a test session in five-trial blocks. Noise level changes occurred after the completion of a given trial block and before the first trial of the next block. With the exception of quiet trials, noise was played constantly throughout a session (i.e., there were no silent intervals between noise level changes) regardless of whether the animal was vocalizing or not. The exact order of levels presented was randomly assigned prior to the start of each session.

All vocalizations produced during the experiment were stored digitally and analyzed off-line using a custom MATLAB signal analysis program. Analysis involved a two-step process in which calls were first sorted by noise level across sessions followed by an acoustic analysis of calls within each noise level. Acoustic measures included average peak frequency, average amplitude, duration, similarity to the template, and 3 dB bandwidth.

C. Results and discussion of experiment 1

Figure 1 shows the increase in call amplitude and call frequency as a function of noise level. The mean level of all calls produced by the birds significantly increased by 7.8 dB SPL as the level of the noise feedback increased from 40 to 90 dB SPL (one-way repeated measures analysis of variance [henceforth RM ANOVA]; $F[11,649]=64.2$, $p<0.01$). Average peak frequency increased by about 84 Hz across the same range of noise levels (one-way RM ANOVA; $F[11,649]=9.75$, $p<0.01$). Call duration also significantly increased with increasing noise level (one-way RM ANOVA; $F[11,649]=4.19$, $p<0.01$). These results show that budgerigars increase fundamental frequency and call length in response to increases in ambient noise levels. Humans and other primates also show an increase in vocal frequency and syllabic length when producing Lombard speech (Lane and Tranel, 1971; Summers *et al.*, 1988; Brumm *et al.*, 2004; Egnor and Hauser, 2006).

Noise level was significantly and inversely proportional to both the similarity of the vocalization to the call template (one-way RM ANOVA; $F[11,649]=2.42$, $p<0.01$) and 3 dB bandwidth (one-way RM ANOVA; $F[11,649]=4.18$, $p<0.01$). The decrease in correlation values indicates that vocalizations produced in noise also contained structural changes in the call. These changes may be related to the decrease in call bandwidth, which itself may reflect a strategy to increase call detectability in noise. Prior work on the perception of vocalizations in noise by budgerigars and zebra finches has shown that for the same overall level, narrow band vocalizations are more easily detected in noise than wide band vocalizations (Lohr *et al.*, 2003).

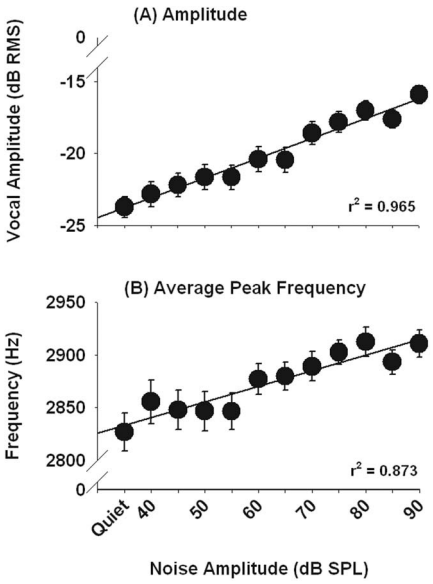


FIG. 1. Increases in amplitude and average peak frequency from calls produced in noise during experiment 1 (Lombard effect). Means and standard errors are plotted.

IV. EXPERIMENT 2: FLETCHER EFFECT

Experiment 2 examined the effect of a level increase in AF on production amplitude. The Fletcher effect, which as far as the authors know has only been reported for humans, is a decrease in vocal amplitude in response to an increase in perceived vocal loudness. As in the previous experiment, birds were trained to produce a specific call in the operant environment. Based on results from human studies and also from experiment 1, the authors hypothesized that (1) vocal amplitude would decrease as the level of the AF increased and (2) there would be concomitant decreases in both average vocal frequency and call duration.

A. Method

1. Subjects

The three birds from experiment 1 were used in this experiment.

B. Procedure

Each bird's stored call template was used as the feedback stimulus and delivered through the earphones while the bird vocalized. The template call was used instead of the bird's actual vocalizations to ensure that the feedback was delivered at specific levels and was not dependent on scaling a bird's own vocal production (which itself was expected to

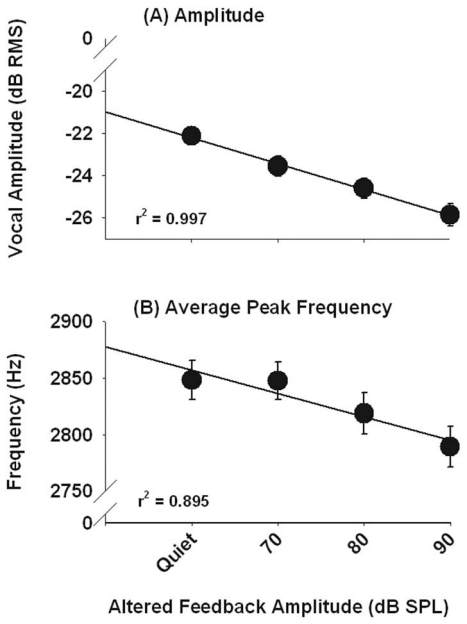


FIG. 2. Decreases in amplitude and average peak frequency from calls produced during experiment 2 (Fletcher effect). Means and standard errors are plotted.

vary in amplitude during a session as the feedback level was changed). Observations from training sessions show that these calls tend to be highly stereotyped under operant control, with a standard deviation under 5 ms in duration and under 3 dB in amplitude within a test session. Thus, the template call functioned satisfactorily as a surrogate for the bird's actual production.

The template was stored in a memory buffer in the TDT RP2.1 and delivered through the earphones when a vocalization was detected at the microphone. The authors used four different feedback levels (i.e., quiet, 70, 80, and 90 dB SPL, *A*-weighting, fast rms) that were measured using a Larson-Davis model 824 sound level meter. These values were chosen based on the level at which the birds typically vocalized in the operant environment, which was about 70 dB SPL. Birds likely also heard their own vocalizations in each of these conditions through both air [due to the lack of complete occlusion by the headphones (see Sec. IID above)] and bone conduction.

Birds were run in two sessions of 40 trials each. A trial was defined by a single vocalization produced in response to an illuminated LED. Feedback levels were presented in five-trial blocks and each trial block was tested twice per session. The order of the trial blocks was randomly assigned prior to each session. All vocalizations were stored digitally and analyzed off-line using a MATLAB signal analysis program. Analysis involved a two-step process in which calls were first sorted by feedback level followed by an acoustic analysis of calls within each noise level. Acoustic measures included average peak frequency, average amplitude, duration, similarity to the template, and 3 dB bandwidth.

C. Results and discussion of experiment 2

Figure 2 shows the decrease in call amplitude and call frequency as a function of feedback level. The mean level of all calls produced by the birds significantly decreased by 3.74 dB between the quiet condition and the 90 dB SPL feedback condition (one-way RM ANOVA; $F[3,177]=38.5$, $p<0.01$). Mean call frequency for the three birds across feedback levels significantly decreased by 59.1 Hz across feedback levels (one-way RM ANOVA; $F[3,177]=13.4$, $p<0.01$). Call duration also significantly decreased as feedback level increased (one-way RM ANOVA; $F[3,177]=2.74$, $p<0.05$). There were no significant differences in either the similarity to the template call or 3 dB bandwidth.

The authors also looked at the time course of these changes by comparing the amplitude and frequency of the first two calls produced after the feedback level changed. There were no differences in amplitude ($F[3,6]=1.72$, $p=0.26$) or frequency ($F[3,6]=1.08$, $p=0.43$) across feedback levels on the first trial, but calls were significantly different across feedback levels in both amplitude ($F[3,6]=5.19$, $p<0.05$) and frequency ($F[3,6]=5.12$, $p<0.05$) on the second feedback trial. Thus, budgerigars do not make online adjustments to either the amplitude or frequency of these short contact calls but rather adjust both frequency and amplitude on the subsequent vocalization.

V. EXPERIMENT 3: DELAYED AND ALTERED AUDITORY FEEDBACK

Here the authors tested the effect of altering AF of the bird's own vocalization on vocal production. There were three altered feedback conditions: (1) DAF of the bird's own vocalizations ("DAF" condition), (2) a temporally-reversed version of the birds' template call ("reversed" condition), and (3) another bird's call as the altered feedback stimulus ("other" condition). Based on prior work in humans and songbirds, the authors hypothesized that altered feedback would disrupt normal call production by inducing changes in pitch, duration, and other spectrotemporal aspects of the call.

A. Method

1. Subjects

The three birds from experiments 1 and 2 were used in this experiment.

B. Procedure

Birds were trained with the same methods of the previous two experiments to produce contact calls that had a spectral cross-correlation criterion to the template of at least $r=0.70$. Then, in three conditions, altered feedback was presented through the earphones whenever a vocalization was detected at the microphone. In the DAF condition, the altered feedback stimulus was the bird's own vocalization from the incoming microphone signal delayed in time by 0, 25, 50, 75, or 100 ms. In both the reversed and other conditions, the stimulus presented was either a reversed version of the birds' template call or the contact call of another bird as feedback stimuli, respectively. These last two stimulus types were stored in a memory buffer and presented with a 0 ms delay. The level of the altered feedback stimuli was calibrated to 70

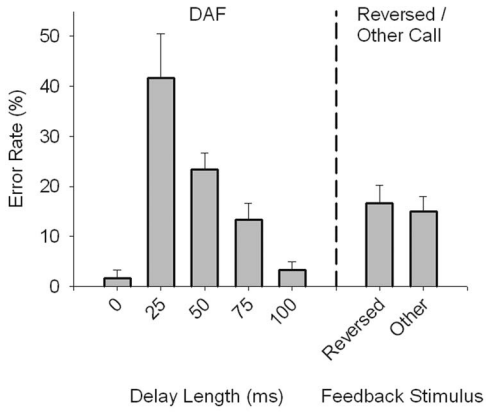


FIG. 3. Error rate varies as a function of feedback type (e.g., DAF delay length—left panel; reversed/other call—right panel). Mean and standard errors for each of the five delay lengths show that birds produced errors on the first call after altered feedback 41.7% of the time in those sessions with feedback delays of 25 ms. Fewer errors were produced as delay length increased. Similarly, errors were produced on the first call after feedback in the reversed and other conditions on 16.7% and 15.0% of trials, respectively.

dB SPL (typical of birds vocalizing in the operant chamber) using a Larson-Davis model 824 sound level meter.

Birds were run in two sessions for each feedback type (e.g., five DAF delay values, reversed, and other) for a total of 14 sessions. Each session was comprised of 70 total trials: 10 altered feedback trials and 60 non-altered feedback trials. The ten altered feedback trials were randomized within each session so that one altered feedback trial was presented for every three to eight non-altered feedback trials. Also, the order of each feedback type was randomized across testing sessions. All birds completed the full 70 trials for all sessions.

Vocalizations were stored digitally and analyzed off-line using a MATLAB signal analysis program. Calls were first sorted by trial type (i.e., pre-altered feedback trial, altered feedback trial, first trial post-altered feedback, etc.) and then acoustic measures—including peak fundamental frequency, amplitude, duration, and similarity to the template—were calculated to compare calls across different trial and stimulus types.

C. Results and discussion of experiment 3

The three birds showed disruptions in vocal behavior in all three altered feedback conditions. The most obvious finding was that birds often produced calls that fell below a spectral cross-correlation to the template of $r=0.70$, which was the minimum value required for reward during training sessions. These calls were labeled errors. Most of these errors (83.0%) occurred within the first two calls after an altered feedback trial and most of those (77.3%) occurred on the first call following altered feedback presentation. No errors occurred during the altered feedback trials.

Figure 3 shows the error rate on the first trial following the altered AF trial across the three conditions. Under the DAF condition, error rates differed as a function of delay length (one-way RM ANOVA; $F[4,8]=24.01$, $p<0.001$). As in humans, there was a maximally disruptive delay that resulted in the most errors, which for budgerigars was 25 ms.

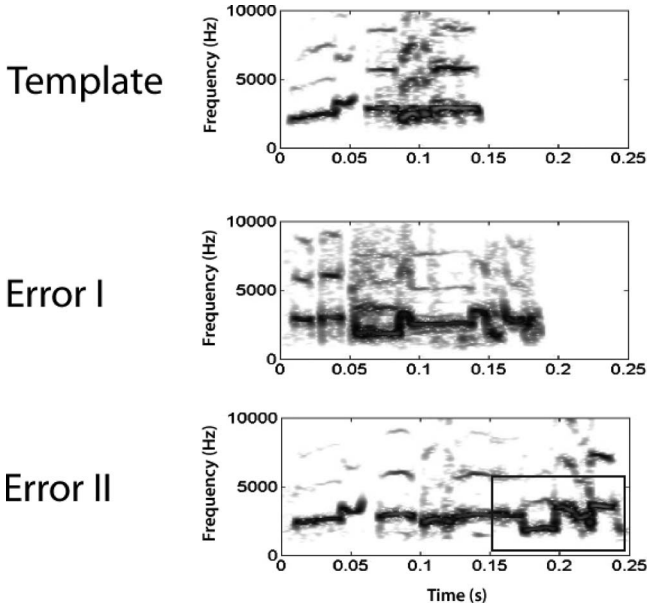


FIG. 4. Example spectrograms from experiment 3 showing one bird's trained template call and an example from each of the two error categories produced by the same bird. The first kind of error (error I) occurred when a subject produced a call-type different from the trained call. Calls in which a bird added or subtracted call elements from its trained call were labeled error II. In the error II example shown, the bird added about 100 ms to the end of its call (boxed area).

Both the reversed and other conditions produced statistically indistinguishable error rates of 16.7% and 15.0%, respectively (paired samples t-test; $t[2]=-2.00$, $p=0.18$), and the average error rate between these two conditions was significantly lower than in the 25 ms DAF condition (paired samples t-test; $t[2]=5.20$, $p<0.05$).

Calls classified as errors fell into two general categories based on visual inspection of spectrographic characteristics. The first type of error occurred when the bird produced a call-type different from the template call (error I). This type of error represented 62.3% of all errors. The second kind of error (error II) was less common (37.7%) and occurred when the bird produced the same call-type as the template call but with new or missing acoustic elements. An example of each of these types of errors is shown in Fig. 4. Overall, there were 180 feedback trials resulting in a total of 88 errors of all types.

VI. GENERAL DISCUSSION

In three experiments the authors show that budgerigar vocalizations are affected by real-time AF. Experiments 1 and 2 showed that these birds exhibit the Lombard and Fletcher effects—increasing vocal amplitude in the presence of background noise and decreasing vocal amplitude when the AF level of their calls is increased. Vocal amplitude increased linearly 1 dB for every 5 dB change in AF loudness. This is a shallower slope than found in humans, where a 1 dB change in vocal amplitude occurs for each 2–3 dB change in AF loudness (see Lane and Tranel, 1971). The difference in slope between humans and budgerigars may be a function of the natural vocal level range of each species. That is,

humans can typically speak across 50+dB range of level (from whisper to shout), whereas budgerigars probably have a much smaller range.

The authors also showed that these amplitude changes are accompanied by changes in contact call frequency, duration, bandwidth, and other acoustic characteristics that are also correlated with vocal effort in speech (e.g., Traunmüller and Eriksson, 2000). In humans, an increase in vocal effort results in increased amplitude, duration, and pitch of speech while a decreasing vocal effort shows the opposite effects. Like humans, birds also produce sound driven by respiratory airflow through a set of vibrating structures in the sound producing organ (e.g., the syrinx in birds and the larynx in humans) (Fletcher and Tarnopolsky, 1999; Larsen and Goller, 1999; 2002). From the pattern of vocal changes occurring during altered feedback trials, budgerigars appear to alter their vocal effort, in a manner analogous to humans, by altering the velocity of the air passing through the syringeal membranes without changing membrane tension (e.g., Heaton *et al.*, 1995; Brittan-Powell *et al.*, 1997). Changes such as increased amplitude and duration and decreased bandwidth all have the practical effect of increasing audibility in noise (Lane and Tranel, 1971; Summers *et al.*, 1988; Lohr *et al.*, 2003).

Experiment 3 also showed that budgerigars are affected by other alterations of AF. Delaying or reversing the vocalizations, or presenting another bird's contact call as AF, resulted in a range of production errors. These errors were similar to those reported in humans and songbirds (e.g., Yates, 1963; Leonardo and Konishi, 1999; Cynx and Von Rad, 2001; Sakata and Brainard, 2006) and included changes in peak frequency, amplitude, and duration. Errors were generally of two types: either production of a different call-type or production of a call with additions or omissions of elements. Interestingly, these errors occurred in subsequent vocalizations but never during the altered feedback trial. This result is consistent with those from experiment 2 (Fletcher effect) showing that amplitude adjustments do not occur in real-time but instead occur on the subsequent call. Taken together, these results on budgerigars show both similarities and differences with analogous behavioral results in humans.

In both humans and birds, the physiological mechanisms underlying altered feedback effects remain obscure. In humans, recent work shows evidence for vocalization-induced suppression of auditory cortex neural activity during ongoing speech (e.g., Houde *et al.*, 2002). Several functional magnetic resonance imaging (fMRI) studies have shown that altered AF, including DAF, activates areas in and around auditory cortex, superior temporal lobe, and planum temporale within 100–130 ms during speech (Hashimoto and Sakai, 2003; Guenther, 2006). Similar patterns of excitation and suppression have been described more recently in nonhuman primates and may result from common mechanisms (e.g., Eliades and Wang, 2008).

In songbirds, vocal errors appear to be processed in forebrain premotor areas. For example, Sakata and Brainard (2006) showed that DAFs of single song syllables provided to Bengalese finches at delays ranging from 40 to 65 ms were most effective in generating vocal errors, although errors occurred with delays as short as 20 ms. Errors did not occur within a syllable, but in subsequent syllables. The authors therefore speculated that the feedback signal is processed in the forebrain premotor nucleus HVC, which is involved in syllable sequencing. Consistent with this hypothesis, a more recent study has shown that HVC contains a population of neurons that are activated by both hearing a song and producing the same song, which could serve an important role in error-correction processes (Prather *et al.*, 2008).

The neural underpinnings of budgerigar vocal feedback control are much less understood than in either humans or songbirds. The budgerigar nucleus NLc, a telencephalic vocal motor region possibly analogous to songbird HVC, responds to auditory input within about 100 ms (Plummer and Striedter, 2000) and projects to striatal structures responsible for learning new contact calls (Striedter, 1994; Brauth *et al.*, 1997). If this nucleus is functionally similar to songbird HVC, it might also contain a population of neurons responsible for comparing actual and expected feedback, and could be responsible for selecting the correct contact call prior to production. The fact that NLc does not receive AF information until 100 ms after vocal onset might explain the bird's inability to use altered feedback for correcting online vocal errors for short sounds since typical budgerigar contact calls are typically 100–150 ms (Farabaugh and Dooling, 1996). It could be used to adjust subsequent vocalizations, however, which is consistent with the behavioral data described here. In addition, if NLc guides selection of the correct circuitry underlying call production, this may also explain why DAF results in erroneous call selection or the production of alternate call-types. Results of the present behavioral experiments are at least consistent with such a function for budgerigar NLc.

The fact that altered AF in budgerigars affects subsequent calls rather than the ongoing vocalization is different from what has been found in humans and songbirds (e.g., Yates, 1963; Leonardo and Konishi, 1999; Cynx and Von Rad, 2001; Sakata and Brainard, 2006). In part, this may have to do with the fact that budgerigar contact calls are so short that feedback mechanisms do not have time to engage. Vocal amplitude adjustments following altered feedback during sustained vowel production in humans does not occur until approximately 150–175 ms of feedback onset (Heinks-Maldonado and Houde, 2005; Bauer *et al.*, 2006). Similarly, DAF effects in humans and songbirds occur at delays of about 200 ms (e.g., Yates, 1963; Howell and Archer, 1984) and 50–100 ms (Cynx and Von Rad, 2001; Sakata and Brainard, 2006), respectively. These are about the duration of a typical budgerigar contact call (about 150 ms) and suggest that the physiological response to altered AF may require a minimum latency greater than the length of a call.

Another possibility is that contact calls are produced ballistically and cannot be modified once initiated. In contrast, zebra finches and tamarins will interrupt their vocal production in structured ways in response to a strobe light (e.g., Cynx, 1990; Miller *et al.*, 2003). The type of errors budgerigars made in experiment 3 did not appear to be examples of truncated contact calls and all errors occurred after

the trial with altered feedback. Instead, budgerigar vocal errors are better described as different, intact call-types. This suggests that incorrect AF might be disrupting selection of the correct motor program sequence that gives rise to the next call-type. In songbirds, onset delays from altered AF are about the average length of song syllables, which are themselves single vocal motor gestures and probably produced ballistically (Cynx, 1990; Riebel and Todt, 1997; Franz and Goller, 2002). While there is evidence that budgerigars learn new calls through a process of recombination and modification of smaller call elements (Farabaugh *et al.*, 1994; Manabe and Dooling, 1997), the present results argue more that the entire call is produced ballistically rather than just the individual elements.

In sum, these results show that AF in budgerigars, as in humans and songbirds, is used to guide future vocal production. The authors measured changes in call amplitude, frequency, and duration that are consistent with the idea that budgerigar vocal production contains mechanisms for overcoming the masking effects of environmental noise. The authors also showed that temporally- and spectrally-misaligned feedback interrupt call production. Our results, in which altered AF affects subsequent calls, differ from both the human and songbird cases in which AF is used to make online adjustments to vocalizations as they are produced. These results could argue for an error-correction mechanism similar to that reported in humans and songbirds but which operates on a time scale greater than the length of a contact call. Alternatively, these results may indicate that different mechanisms are involved in AF-guided vocal production in these two different groups of vocal learners.

ACKNOWLEDGMENTS

The authors thank Elizabeth Brittan-Powell and Peter Marvit for their many helpful comments on the manuscript. They also thank Leah Dickstein for help in performing the experiments. This work was supported by NIH Grant Nos. DC-00198 and DC-04664 to R.J.D. and NIH Grant No. DC-006766 to M.S.O.

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.

Bloodstein, O. (**1995**). *A Handbook on Stuttering* (National Easter Seal Society, Chicago, IL).

Boughman, J. W., and Moss, C. F. (**2003**). "Social sounds: Vocal learning and development of mammal and bird calls," in *Acoustic Communication, Springer Handbook of Auditory Research*, edited by A. M. Simmons, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 138–224.

Brainard, M. S., and Doupe, A. J. (**2000**). "Auditory feedback in learning and maintenance of vocal behaviour," Nat. Rev. Neurosci. **1**, 31–40.

Brauth, S. E., Heaton, J. T., Shea, S. D., Durand, S. E., and Hall, W. S. (**1997**). "Functional anatomy of forebrain vocal control pathways in the budgerigar (Melopsittacus undulatus)," Ann. N.Y. Acad. Sci. **807**, 368–385.

Brittan-Powell, E. F., Dooling, R. J., Larsen, O. N., and Heaton, J. T. (**1997**). "Mechanisms of vocal production in budgerigars (Melopsittacus undulatus)," J. Acoust. Soc. Am. **101**, 578–589.

Brumm, H., and Todt, D. (**2002**). "Noise-dependent song amplitude regulation in a territorial songbird," Anim. Behav. **63**, 891–897.

Brumm, H., Voss, K., Kollmer, I., and Todt, D. (**2004**). "Acoustic communication in noise: Regulation of call characteristics in a New World monkey," J. Exp. Biol. **207**, 443–448.

Chase, R. (**1965**). "An information-flow model of the organization of motor activity I: Transduction, transmission, and central control of sensory information," J. Nerv. Ment. Dis. **140**, 239–251.

Cynx, J. (**1990**). "Experimental-determination of a unit of song production in the zebra finch (*Taeniopygia guttata*)," J. Comp. Psychol. **104**, 3–10.

Cynx, J., Lewis, R., Tavel, B., and Tse, H. (**1998**). "Amplitude regulation of vocalizations in noise by a songbird, Taeniopygia guttata," Anim. Behav. **56**, 107–113.

Cynx, J., and Von Rad, U. (**2001**). "Immediate and transitory effects of delayed auditory feedback on bird song production," Anim. Behav. **62**, 305–312.

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

Draegert, G. L. (**1951**). "Relationships between voice variables and speech intelligibility in high level noise," Speech Monogr. **18**, 272–278.

Dreher, J. J., and O'Neill, J. J. (**1958**). "Effects of ambient noise on speaker intelligibility of words and phrases," Laryngoscope **68**, 539–548.

Egnor, S. E. R., and Hauser, M. D. (**2006**). "Noise-induced vocal modulation in cotton-top tamarins (Saguinus oedipus)," Am. J. Primatol. **68**, 1183–1190.

Eliades, S. J., and Wang, X. (**2008**). "Neural substrates of vocalization feedback monitoring in primate auditory cortex," Nature (London) **453**, 1102–1106.

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**, 333–346.

Farabaugh, S., and Dooling, R. J. (**1996**). "Acoustic communication in parrots: Laboratory and field studies of budgerigars, Melopsittacus undulates," in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell University Press, Ithaca, NY), pp. 97–118.

Farabaugh, S. M., Dent, M. L., and Dooling, R. J. (**1998**). "Hearing and vocalizations in native Australian budgerigars (Melopsittacus undulatus)," J. Comp. Psychol. **112**, 74–81.

Farabaugh, S. M., Linzenbold, A., and Dooling, R. J. (**1994**). "Vocal plasticity in budgerigars (*Melopsittacus undulatus*): Evidence for social factors in the learning of contact calls," J. Comp. Psychol. **108**, 81–92.

Fletcher, H., Raff, G. M., and Parmley, F. (**1918**). "Study of the effects of different sidetones in the telephone set," Report No. 19412, Western Electric Company, New York, N.Y.

Fletcher, N. H., and Tarnopolsky, A. (**1999**). "Acoustics of the avian vocal tract," J. Acoust. Soc. Am. **105**, 35–49.

Franz, M., and Goller, F. (**2002**). "Respiratory units of motor production and song imitation in the zebra finch," J. Neurobiol. **51**, 129–141.

Guenther, F. H. (**2006**). "Cortical interactions underlying the production of speech sounds," J. Commun. Disord. **39**, 350–365.

Hanley, T. D., and Steer, M. D. (**1949**). "Effect of level of distracting noise upon speaking rate, duration and intensity," J. Speech Hear Disord. **14**, 363–368.

Hashimoto, Y., and Sakai, K. (**2003**). "Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: An fMRI study," Hum. Brain Mapp **20**, 22–28.

Heaton, J. T., Farabaugh, S. M., and Brauth, S. E. (**1995**). "Effect of syringeal denervation in the budgerigar (Melopsittacus undulatus): The role of the syrinx in call production," Neurobiol. Learn Mem. **64**, 68–82.

Heinks-Maldonado, T. H., and Houde, J. F. (**2005**). "Compensatory responses to brief perturbations of speech amplitude," ARLO **6**, 131–137.

Houde, J. F., Nagarajan, S. S., Sekihara, K., and Merzenich, M. M. (**2002**). "Modulation of the auditory cortex during speech: An MEG study," J. Cogn Neurosci. **14**, 1125–1138.

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

Janik, V. M., and Slater, P. J. B. (**1997**). "Vocal learning in mammals," in *Advances in the Study of Behaviour*, edited by P. J. B. Slater, J. S. Rosenblatt, C. T. Snowdon, and H. Milinski (Academic, New York), pp. 59–99.

Janik, V. M., and Slater, P. J. B. (**2000**). "The different roles of social learning in vocal communication," Anim. Behav. **60**, 1–11.

Kobayasi, K. I., and Okanoya, K. (**2003**). "Context-dependent song amplitude control in Bengalese finches," NeuroReport **14**, 521–524.

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

Larsen, O. N., and Goller, F. (**1999**). "Role of syringeal vibrations in bird vocalizations," Proc. R. Soc. London, Ser. B **266**, 1609–1615.
Larsen, O. N., and Goller, F. (**2002**). "Direct observation of syringeal muscle functions in songbirds and a parrot," J. Exp. Biol. **205**, 25–35.
Lee, B. S. (**1950**). "Effects of delayed speech feedback," J. Acoust. Soc. Am. **22**, 824–826.
Leonardo, A., and Konishi, M. (**1999**). "Decrystallization of adult birdsong by perturbation of auditory feedback," Nature (London) **399**, 466–470.
Lohr, B., Wright, T. F., and Dooling, R. J. (**2003**). "Detection and discrimination of natural calls in masking noise by birds: Estimating the active space signal," Anim. Behav. **65**, 763–777.
Lombard, E. (**1911**). "Le signe de l'élévation de la voix (The characteristics of the elevation of the voice)," Ann. Maladies Oreille, Larynx, Nez, Pharynx (Annals of the Diseases of the Ear, Larynx, Nose, and Pharynx) **37**, 101–119.
Manabe, K., and Dooling, R. J. (**1997**). "Control of vocal production in budgerigars (Melopsittacus undulatus): Selective reinforcement, call differentiation, and stimulus control," Behav. Processes **41**, 117–132.
Manabe, K., Sadr, E. I., and Dooling, R. J. (**1998**). "Control of vocal intensity in budgerigars (*Melopsittacus undulatus*): Differential reinforcement of vocal intensity and the Lombard effect," J. Acoust. Soc. Am. **103**, 1190–1198.
Miller, C. T., Flusberg, S., and Hauser, M. D. (**2003**). "Interruptibility of long call production in tamarins: Implications for vocal control," J. Exp. Biol. **206**, 2629–2639.
Plummer, T. K., and Striedter, G. F. (**2000**). "Auditory responses in the vocal motor system of budgerigars," J. Neurobiol. **42**, 79–94.
Potash, L. M. (**1972**). "Noice-induced changes in calls of the Japanese quail," Psychonomic Sci. **26**, 252–254.
Prather, J. F., Peters, S., Nowicki, S., and Mooney, R. (**2008**). "Precise auditory-vocal mirroring in neurons for learned vocal communication," Nature (London) **451**, 305–310.
Pytte, C. L., Rusch, K. M., and Ficken, M. S. (**2003**). "Regulation of vocal amplitude by the blue-throated hummingbird, *Lampornis clemenciae*," Anim. Behav. **66**, 703–710.
Riebel, K., and Todt, D. (**1997**). "Light flash stimulation alters the nightingale's singing style: Implications for song control mechanisms," Behaviour **134**, 789–808.
Sakata, J. T., and Brainard, M. S. (**2006**). "Real-time contributions of auditory feedback to avian vocal motor control," J. Neurosci. **26**, 9619–9628.
Siegel, G. M., and Pick, H. L. (**1974**). "Auditory feedback in the regulation of voice," J. Acoust. Soc. Am. **56**, 1618–1624.
Sinnott, J. M., Stebbins, W. C., and Moody, D. B. (**1975**). "Regulation of voice amplitude by the monkey," J. Acoust. Soc. Am. **58**, 412–414.
Striedter, G. F. (**1994**). "The vocal control pathways in budgerigars differ from those in songbirds," J. Comp. Neurol. **343**, 35–56.
Summers, W. V., 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.
Traunmüller, H., and Eriksson, E. (**2000**). "Acoustic effects of variation in vocal effort by men, women, and children," J. Acoust. Soc. Am. **107**, 3438–3451.
Yates, A. J. (**1963**). "Delayed auditory feedback," Psychol. Bull. **60**, 213–232.

Erratum: Beamwidth control and angular target detection in an echolocating dolphin (Tursiops truncatus) [J. Acoust. Soc. Am. 124, 3324–3332 (2008)]

Patrick W. Moore
SPAWARSYSCEN-San Diego, Code 2351, 53560 Hull Street, San Diego, California 92152-5001

Lois A. Dankiewicz
Science Applications International Corporation, 4065 Hancock Street, Mailstop Q1–A, San Diego, California 92110

Dorian S. Houser
BIOMIMETICA, 7951 Shantung Drive, Santee, California 92071

(Received 28 May 2009; accepted 3 June 2009)
[DOI: 10.1121/1.3158940]

PACS number(s): 43.80.Ka, 43.80.Jz, 43.10.Vx

On page 3326, second paragraph, line 10 the spacing of the hydrophones in the array should read with a spacing of 27 cm horizontal and 18 cm vertical arc separation between each hydrophone.

 0001-4966/2009/126(2)/920/1/$25.00

ACOUSTICAL NEWS

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

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

Preliminary Notice: 158th Meeting of the Acoustical Society of America

The 158th Meeting of the Acoustical Society of America will be held Monday through Friday, 26-30 October 2009 at the Hyatt Regency San Antonio in San Antonio, Texas. Information about the meeting also appears on the ASA Home Page at ⟨http://asa.aip.org/meetings.html⟩.

Charles E. Schmid
Executive Director

Technical Program

The technical program will consist of lecture and poster sessions. Technical sessions will be scheduled Monday through Friday, 26-30 October 2009.

Special Sessions

ACOUSTICAL OCEANOGRAPHY (AO)

Acoustic measurements of water column scatterers using multibeam sonar
Multibeam sonar measurements of water column scatterers (e.g., fish, bubbles, suspended sediments, internal waves, and fine structure)
Acoustics and ocean acidity
Sensitivity of low frequency sound absorption to ocean pH can in principle be exploited to measure ocean acidity. Papers on possible measurement approaches, ocean acidity and its trends, and potential effects on ambient noise are encouraged
Session in honor of Stan Flatté
(Joint with Underwater Acoustics)
Stan Flatté's contributions to understanding acoustic fluctuations in the ocean and the basic science of wave propagation though random media. Papers on theoretical and observational aspects, and signal processing and ocean inversion implications are encouraged

ANIMAL BIOACOUSTICS (AB)

Acoustical monitoring of animals to track climate, biotic, and environmental changes
(Joint with Acoustical Oceanography)
How assessment of animal populations using acoustics can help assess environmental changes
Acoustical surveys, sampling, and population assessment of animals
How acoustics can be used to survey, sample, and assess animal populations
Emotion-related mechanisms of primate vocalizations
Research on mechanisms of primate vocalization, especially in emotional context
Natural soundscapes and auditory scene analysis by animals
Descriptions and analysis of natural "auditory scenes" used by animals for orientation and perceptual mechanisms

ARCHITECTURAL ACOUSTICS (AA)

Absorption, scattering, and diffusion: Myths, facts, misconceptions
Information and testing methods regarding absorption, diffusion and scattering. What are the differences? What do we know that is true? What is myth and misinformation? How do we use the information?
Acoustic test facilities: Design, qualification and testing
(Joint with Noise and ASA Committee on Standards)
Acoustics of hemi-anechoic chambers and reverberation chambers including design, qualifications, and testing methodologies
Acoustics and theater consulting: A special relationship
(Joint with Noise and Musical Acoustics)
Panel discussion with theater consultants discussing the theater consultant's joys and challenges in working with acoustics consultants
Architectural acoustics program serving consultants and professionals
(Joint with Education in Acoustics)
Overview of current program content and paradigms of approach followed by a panel discussion
Biggest mistakes: Lessons for practitioners, researchers, and young designers
Opportunity to learn not from others' triumphs but from their hard lessons
Classroom acoustics: An update
(Joint with Noise and ASA Committee on Standards)
Review of active progress in classroom acoustics
Classroom acoustics: New design approaches both successes and failures
(Joint with ASA Committee on Standards, Noise, and Speech Communication)
New design approaches to improve the classroom environment with the best interests of the students in mind
ETS-Lindgren acoustic test laboratories and factory: Papers and tour
(Joint with Noise)
A tour of the National Voluntary Laboratory Accreditaton Program (NVLAP)-accredited ETS-Lindgren factory and acoustic research laboratory which includes a large double-wall hemianechoic chamber and reverberation chamber suite
Seismic restraint, structural issues and building acoustics
(Joint with Noise and Structural Acoustics and Vibration)
How to achieve desired acoustics while meeting seismic and other structural requirements
Spaces for experimental media and spatial sound reproduction
(Joint with Engineering Acoustics)
Exploration of acoustic design challenges in constructing spaces for complex multimedia presentations and surround sound for large venues

BIOMEDICAL ULTRASOUND/BIORESPONSE TO VIBRATION (BB)

Molecular imaging with ultrasound
(Joint with Physical Acoustics)
Development of ultrasound contrast agents and ultrasound imaging systems for molecular imaging and sensing
Multiple scattering of waves: From theory to application
(Joint with Physical Acoustics and Underwater Acoustics)
Focus on all fields of acoustics where multiple scattering phenomena play a role, at various frequencies. Different applications (tissue and materials characterization, backscatter from bubble clouds and schools of fish, etc.) will be considered. Focus will be on potential shifts in resonances, changes in backscatter amplitude, and separating coda in the time domain

ENGINEERING ACOUSTICS (EA)

Acoustic measurement and models for sensors and arrays
(Joint with Signal Processing in Acoustics)
Recent advancement of theoretical or numerical model and devices, in acoustic sensor/array measurement and calibration
Magnetostriction in Galfenol and its applications
Recent materials advances in Galfenol and applications in acoustics and vibration
Geophysical acoustics
(Joint with Signal Processing in Acoustics)
Advances in acoustics for geophysical exploration: seismic technology, borehole acoustics, design and modeling of acoustic transducers for down-

 0001-4966/2009/126(2)/921/6/$25.00

hole and surface applications, crosswell acoustics, acoustics of rock-porous media, and advances in signal processing and data analysis for acoustic well logging and seismic applications

EDUCATION IN ACOUSTICS (ED)

Hands-on experiments for high school students
Experiments for high school students
Historical teaching and research acoustic apparatus
Descriptions of both teaching and research apparatus that is antique or historical and no longer in use, except possibly in museums
"Project Listen Up"
(Joint with ASA Student Council)
Descriptions of acoustic demonstrations, laboratory experiments or discovery activities for learners of all ages. Apparatus may be shown but the talks should focus on concepts, explanations, diagrams and drawings with an emphasis on careful scientific approach

MUSICAL ACOUSTICS (MU)

Acoustics of free-reed instruments
Wide range of studies on free-reed instruments and playing techniques with a focus on the accordion and concertina
Acoustics of percussion instruments
Wide range of studies on percussion instruments and playing techniques
Musical ability: Innate and environmental influences
Innate and environmental influences on musical ability
Psychoacoustic response to musical instruments
Perception and acoustic responses to the enjoyment of musical instruments

NOISE (NS)

Advancements in noise control in Latin America
New developments in noise control standards and techniques
Advancements in soundscape application and standards
(Joint with ASA Committee on Standards)
Soundscape concepts and their application in city management
Annoyance of vehicle exterior noise
Measurement, analysis, and prediction
Sustainable building systems and acoustic performance
(Joint with Architectural Acoustics)
How to achieve environmentally conscious mechanical/building systems while maintaining a low-noise profile

PHYSICAL ACOUSTICS (PA)

A man for all seasons: Tribute to Robert T. Beyer
(Joint with Engineering Acoustics)
Professor Beyer's career expressed as a portrait as seen through the eyes of his students and colleagues
40th anniversary of the Khokhlov/Zabolotskaya/Kuznetsov (KZK) equation
(Joint with Biomedical Ultrasound/Bioresponse to Vibration)
Past and present studies of nonlinear sound beams based on the Khokhlov/Zabolotskaya/Kuznetsov (KZK) equation, including mathematical properties, methods of solution, and applications of this equation
Light and sound in science, engineering, and medicine
(Joint with Biomedical Ultrasound/Bioresponse to Vibration and Engineering Acoustics)
Synergistic use of light and sound for sensing, imaging, and inducing physical effects

SIGNAL PROCESSING IN ACOUSTICS (SP)

Categorization of animal acoustic signals
(Joint with Animal Bioacoustics, Acoustical Oceanography and Underwater Acoustics)
Fundamental signal processing methodologies that could be applied to categorization of animal acoustic signals; contributions expected on classification and detection, along with Bayesian methods for recognition and estimation problems
Time-frequency theory and applications
Theory and applications of various time-frequency methods used in acoustic signal processing. This includes, Wigner distributions, Zak transforms, fractional Fourier transforms and other related methods
Kalman and particle filters in acoustics
Theory and the importance and applicability of Kalman and particle filters to all areas of acoustics

SPEECH COMMUNICATION (SC)

Advances in speech synthesis
Recent developments in speech and voice synthesis with emphasis on theoretical motivations and variations across applications
Fluid-structure interaction in voice production: Experiments and modeling
Experimental and numerical studies of the glottal flow, vocal fold vibration, and their interaction
Perceptual learning and adaptation in speech
Recent research on how listeners learn to adapt to variability in speech including foreign-accented, degraded, and disordered speech

STRUCTURAL ACOUSTICS AND VIBRATION (SA)

Emerging applications of structural acoustics in energy and power generation
Modeling and prediction of vibro-acoustic characteristics of power generation systems
Structural acoustics and vibrations in nano-materials
(Joint with Physical Acoustics and Engineering Acoustics)
Special dynamic properties, acoustics, and vibration absorption in nano-materials
Structural acoustics for poroelastic materials
(Joint with Physical Acoustics)
Measurement and prediction of transmission losses through poroelastic materials
Structural acoustics in biological media
(Joint with Biomedical Ultrasound/Bioresponse to Vibration)
Sound transmission through cochlea, hearing devices, and other biological media

UNDERWATER ACOUSTICS (UW)

Progress in modeling complex and dynamic acoustic and oceanographic characteristics of continental shelves and slopes
(Joint with Acoustical Oceanography)
Characterization of the environmental variability present on continental slopes and shelves and its effects on acoustic propagation
Reverberation measurements and modeling
Reverberation measurements; modeling papers are encouraged in both intensity and pressure time series

OTHER TECHNICAL EVENTS

Hot Topics

A "Hot Topics" session sponsored by the Tutorials Committee will cover the fields of Acoustical Oceanography, Biomedical Ultrasound/Bioresponse to Vibration and Noise.

Open Meetings of Technical Committees

Technical Committees will hold open meetings on Tuesday, Wednesday, and Thursday evenings. These are working, collegial meetings. Much of the work of the Society is accomplished by actions that originate and are taken in these meetings including proposals for special sessions, workshops and technical initiatives. All meeting participants are cordially invited to attend these meetings and to participate actively in the discussions.

Online Meeting Papers

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

Those interested in obtaining copies of submitted papers for this meeting may access the service at anytime. No password is needed. The url is ⟨http://scitation.aip.org/asameetingpapers/⟩.

Proceedings of Meetings on Acoustics

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

Meeting Program

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

TUTORIAL LECTURE ON PODCASTING DEMYSTIFIED: FROM CONCEPT TO PRODUCTION FOR EVERY BUDGET

A tutorial presentation on "Podcasting DeMystified: From concept to production for every budget" will be given by Kathleen P. King of Fordham University and The Teachers' Podcast ⟨www.teacherspodcast.org⟩ on Monday, 26 October 2009 at 7:00 p.m.

Podcasting offers an inexpensive and yet powerful way to reach people around the globe with your message in audio or video format. Leveraging our global societies' access of 24/7 web-based platforms to meet their information and learning needs, provides a vibrant distribution engine waiting for people of all disciplines to create and post their work. This tutorial will provide an overview of the basic "how-to's" of planning, producing and hosting your podcast episodes and series. Podcasting possibilities may include redistributing content, and events, supplementing learning experiences, extending the traditional classroom and integrating new opportunities. Intellectual property, copyright issues, sound production, equipment options for every budget, free software for recording, editing and production, and resources will be discussed as well. Dr. King has hosted and produced over 8 series, reached over 6 million people through her podcasting work, written 2 books and presented many seminars on this, other digital media and distance learning topics. Utilizing this experience, Dr. King will provide insight in the many different formats, audiences, hosting features, and applications of podcasting. As examples of applications, consider that podcasting can be thought of as teacher-created, student-created or already existing content. We will discuss how you might use these different forms for different valuable purposes in teaching and learning settings.

Lecture notes will be available at the meeting in limited supply. Those who register by 5 October are guaranteed receipt of a set of notes. To partially defray the cost of the lecture a registration fee is charged. The fee is $15.00 USD for registration received by 20 October and $25.00 USD at the meeting. The fee for students with current ID cards is $7.00 USD for registration received by 5 October and $12.00 USD at the meeting. Register online at ⟨http://asa.aip.org⟩ or use the use the registration form in the printed call for papers.

SHORT COURSE ON ACOUSTIC IMAGING APPLICATIONS: SEISMIC, UNDERWATER, BIOMEDICAL, AND SPEECH

Acoustic imaging is both a tool that continues to be used and an emerging area in many disciplines. This short course will introduce how acoustic imaging is utilized on the forefront of science in four different acoustic disciplines: seismic imaging, underwater acoustic imaging, biomedical ultrasound imaging, and ultrasound imaging of speech.

The short course will be taught by a team of instructors who cover a wide range of expertise in acoustic imaging. Dr. Max Deffenbaugh is a research geophysicist with ExxonMobil Research & Engineering Company, Annandale, NJ, who works on seismic signal processing and the estimation of rock properties from seismic data. Dr. Jules Jaffe is a Research Oceanographer at the Scripps Institution of Oceanography and a fellow of the ASA who has worked in the field for 25 years. Dr. Christy Holland is an Associate Professor in the Department of Biomedical Engineering at the University of Cincinnati. She is a fellow of both the ASA and the American Institute of Ultrasound in Medicine, and is currently editor of Ultrasound in Medicine and Biology, the official Journal of the World Federation for Ultrasound in Medicine and Biology. Dr. Maureen Stone is a Professor at the University of Maryland Dental School and a fellow of the ASA. She has used ultrasound to study normal and pathological tongue motion for almost 30 years.

The course schedule is Sunday, 25 October, 1:00 to 5:00 p.m. and Monday, 26 October, 8:30 a.m. to 12:30 p.m.

The registration fee is $250.00 USD and covers attendance, instructional materials and coffee breaks. The number of attendees will be limited so please register early to avoid disappointment. Only those who have registered by 5 October will be guaranteed receipt of instructional materials. There will be a $50.00 USD discount for registration made prior to 20 October. Full refunds will be made for cancellations prior to 20 October. Any cancellation after 5 October will be charged a $25.00 USD processing fee. Register online at ⟨http://asa.aip.org⟩ or use the use the registration form in the printed call for papers.

SPECIAL MEETING FEATURES

Student Transportation Subsidies

A student transportation subsidies fund has been established to provide limited funds to students to partially defray transportation expenses to meetings. Students presenting papers who propose to travel in groups using economical ground transportation will be given first priority to receive subsidies, although these conditions are not mandatory. No reimbursement is intended for the cost of food or housing. The amount granted each student depends on the number of requests received. To apply for a subsidy, submit a proposal (e-mail preferred) to be received by 14 September to: Jolene Ehl, ASA, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, Tel: 516-576-2359, Fax: 516-576-2377, E-mail: jehl@aip.org. The proposal should include your status as a student; whether you have submitted an abstract; whether you are a member of ASA; method of travel; if traveling by auto; whether you will travel alone or with other students; names of those traveling with you; and approximate cost of transportation.

Young Investigator Travel Grant

The Committee on Women in Acoustics (WIA) is sponsoring a Young Investigator Travel Grant to help with travel costs associated with presenting a paper at the San Antonio meeting. Young professionals who have completed their doctorate in the past five years are eligible to apply if they plan to present a paper at the San Antonio meeting, are not currently students, and have not previously received the award. Each award will be of the order of $300 with three awards anticipated. Awards will be presented by check at the WIA luncheon at the meeting. Both men and women may apply. Applicants should submit a request for support, a copy of the abstract for their presentation at the meeting, and a current resume/vita which includes information on their involvement in the field of acoustics and in the ASA. Submission by e-mail is preferred to Dr. Sarah Hargus Ferguson at safergus@ku.edu. Deadline for receipt of applications is 15 September.

Students Meet Members for Lunch

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

sign-up deadline is ten days before the start of the meeting, but an earlier sign-up is strongly encouraged. Each participant pays for his/her own meal.

Plenary Session, Awards Ceremony, Fellows Luncheon and Social Events

Buffet socials with cash bar will be held on Tuesday and Thursday evenings. The ASA Plenary session will be held on Wednesday afternoon, 28 October, at the Hyatt Regency where Society awards will be presented and recognition of newly-elected Fellows will be announced.

A Fellows Luncheon will be held on Thursday, 29 October, at 12:00 noon at the Hyatt Regency. This luncheon is open to all attendees and their guests. Register online at ⟨http://asa.aip.org⟩ or use the registration form in the printed call for papers.

Women in Acoustics Luncheon

The Women in Acoustics luncheon will be held on Wednesday, 28 October. Those who wish to attend this luncheon must register using the form on page 15 or online at ⟨http://asa.aip.org⟩. The fee is $20 (students $10) for pre-registration by 5 October and $25 (students $10) at the meeting. Register online at ⟨http://asa.aip.org⟩ or use the use the registration form in the printed call for papers.

TRANSPORTATION AND HOTEL ACCOMMODATIONS

Air Transportation

San Antonio International Airport (SAT) is located on the north side of San Antonio at the intersection of Interstate Highway Loop 410 and U.S. Highway 281. It is approximately 15 minutes from the central business district as well as Interstate Highway 35 to the east and Interstate Highway 10 to the west. SAT is served by the following airlines: Terminal 1: Aeromexico, Aerolitoral, Delta, Frontier, Mexicana, Midwest, Southwest, Skywest; Terminal 2: American, Continental, and U.S. Air.

Attendees arriving via commercial transportation are encouraged to use public transportation to travel to the Hyatt (rather than car rental) as driving in San Antonio center city can be challenging, parking can be costly and may be limited.

Ground Transportation

There are a variety of ground transportation options between San Antonio International and surrounding metropolitan area destinations. Transportation information is available from the booths located curbside in front of both Terminals 1 and 2. Uniformed transportation agents can describe offered services and provide rates to various destinations. For additional ground transportation information please call (210) 207-3411.

Shared Ride Shuttle: A shuttle service is available from the San Anonio International Airport to downtown hotels for $18.00 per person one-way, or $32.00 round-trip. For additional information visit ⟨http://www.saairportshuttle.com/⟩.

Taxicabs: Taxicabs are available at the lower level curbside, outside of baggage claim, at Terminal 1 and at the front curbside of Terminal 2 just outside baggage claim. The cost from the airport to the Hyatt Regency is approximately $20.00 to $23.00 USD. For additional information visit ⟨http://www.sanantonio.gov/aviation/taxicabs.asp⟩.

Automobile Rental: San Antonio International Airport is served by nine major car rental companies which have counterspaces in each of the Terminals. For additional information visit: ⟨http://www.sanantonio.gov/aviation/carrentals.asp⟩.

Parking at the Hyatt

Hyatt parking is located directly across the street at the Central Parking System garage. They offer self-parking with in/out privileges for $23.00 a day, plus tax and valet parking with in/out privileges for $27.00 a day, plus tax.

Room Sharing

ASA will compile a list of those who wish to share a hotel room and its cost. To be listed, send your name, telephone number, e-mail address, gender, smoker or nonsmoker preference, not later than 18 September to the Acoustical Society of America, preferably by e-mail: asa@aip.org or by postal mail to Acoustical Society of America, Attn.: Room Sharing, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502. The responsibility for completing any arrangements for room sharing rests solely with the participating individuals.

Weather

With 300 days of sunshine a year, San Antonio is an ideal destination year round. Warm summers and mild winters are complemented by a very pleasant spring and fall. The average annual precipitation is just 29.05 inches. In October the daily average temperature rages from 40–87 F, so come prepared for cool weather as well as warm.

HOTEL RESERVATION INFORMATION

A block of guest rooms at discounted rates has been reserved for meeting participants at the Hyatt Regency San Antonio. **Early reservations are strongly recommended**. Note that the special ASA meeting rates are not guaranteed after **3 October 2009**. You must mention the Acoustical Society of America when making your reservations to obtain the special ASA meeting rates.

Hyatt Regency San Antonio

The Hyatt Regency San Antonio is located on the Riverwalk in downtown San Antonio. The Riverwalk is the heart of the San Antonio entertainment district and is a park-like area that borders the San Antonio River as it meanders through the downtown area. The hotel is situated such that rooms have views of the historic Alamo, the Riverwalk and downtown San Antonio.

The Hyatt comprises 632 extra large contemporary guest rooms with Hyatt Grand Beds™, iHome stereo with iPod dock, generous work desk, and Wi-Fi.

Hyatt Regency San Antonio
123 Losoya Street
San Antonio, TX 78205
Tel: (210) 222-1234
Toll-Free: 1-800-233-1234
http://sanantonioregency.hyatt.com/hyatt/hotels/

Online Reservations

http://www.sanantonioregency.hyatt.com/groupbooking/satrsasa72009

Room Rate

Single/Double Occupancy: $189.00
Triple Occupying: $214.00
Quadruple Occupancy: $239.00
All rooms are subject to 16.75% tax
Reservation cut-off date: 3 October 2009

Assistive Listening Devices

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

Child Care

Information about child care services will be provided on the ASA website.

Accompanying Persons Program

Spouses and other visitors are welcome at the San Antonio meeting. The registration fee for accompanying persons is $50.00 for preregistration by 5 October and $75.00 at the meeting. A hospitality room for accompanying persons will be open at the Hyatt Regency from 8:00 a.m. to 10:30 a.m., Monday through Friday. Please check the ASA website at ⟨http://asa.aip.org/meetings.html⟩ for updates about the accompanying persons program.

REGISTRATION INFORMATION

The registration desk at the meeting will open on Monday, 26 October, at the Hyatt Regency. Register online at ⟨http://asa.aip.org⟩ or use the form in the printed call for papers. **If your registration is not received at the ASA headquarters by 5 October you must register on-site**.

Registration fees are as follows:

Category	Preregistration by 5 October	Onsite Registration
Acoustical Society Members	$385	$435
Acoustical Society Members One-Day Attendance*	$195	$245
Nonmembers	$435	$485
Nonmembers One-Day Attendance*	$220	$270
Nonmember Invited Speakers One-Day Attendance*	Fee waived	Fee waived
Nonmember Invited Speakers (Includes one-year ASA membership upon completion of an application)	$110	$110
ASA Early Career Associate or Full Members (For ASA members who transferred from ASA student member status in 2007, 2008, or 2009)	$195	$220
ASA Student Members (with current ID cards)	Fee waived	$25
Nonmember Students (with current ID cards)	$50	$60
Emeritus members of ASA (Emeritus status pre-approved by ASA)	$50	$75
Accompanying Persons (Spouses and other registrants who will not participate in the technical sessions)	$50	$75

Nonmembers who simultaneously apply for Associate Membership in the Acoustical Society of America will be given a $50 discount off their dues payment for the first year (2010) of membership. Invited speakers who are members of the Acoustical Society of America are expected to pay the registration fee, but **nonmember invited speakers** may register for one-day only without charge. A nonmember invited speaker who pays the full-week registration fee, will be given one free year of membership upon completion of an ASA application form.

If you register as a Nonmember Student and complete a membership application, your first year's membership dues (2010) are waived.

*One-day registration is for participants who will attend the meeting for only one day. If you will be at the meeting for more than one day either presenting a paper and/or attending sessions, you must register and pay the full registration fee.

NOTE: A $25 PROCESSING FEE WILL BE CHARGED TO THOSE WHO WISH TO CANCEL THEIR REGISTRATION AFTER 5 OCTOBER.

Calendar of Meetings and Congresses

2009

12–16 August, Jyväskylä, Finland. **7th Triennual Conference of the European Society for Cognitive Science of Music (ESCOM 2009).** Web: http//www.fyu.fi/hum/laitokset/musikki/en/escom2009

23–28 August, Ottawa, Ont. Canada. **Inter-noise 2009.** Web: http://www.internoise2009.com

23–27 Seattle, Washington, USA. **11th International Conference on Music Perception and Cognition.** Web: TBA

06–10 September, Brighton, UK. **InterSpeech 2009 Conference.** Web: http://www.interspeech2009.org

07–11 September, Dresden, Germany. **9th International Conference on Theoretical and Computational Acoustics.** Web: http://ictca2009.com

14–18 September, Kyoto, Japan. **5th Animal Sonar Symposium.** Web: http://cse.fra.affrc.go.jp/akamatsu/AnimalSonar.html

15–17 September, Koriyama, Japan. **Autumn Meeting of the Acoustical Society of Japan.** Web: http://www.asj.gr.jp/index-en.html

19–23 September, Rome, Italy. **IEEE 2009 Ultrasonics Symposium.** E-mail: pappalar@uniroma3.it

21–23 September, Beijing, China. **Western Pacific Acoustics Conference (WESPAC).** Web: http://www.wespacx.org

23–25 September, Xi'an, China. **Pacific Rim Underwater Acoustics Conference (PRUAC).** E-mail: lfh@mail.ioa.ac.cn

23–25 September, Cádiz, Spain. **TECNIACUSTICA'09.** Web: http://www.-sea-acustica.es

05–07 October, Tallinn, Estonia. **International Conference on Complexity of Nonlinear Waves.** Web: http://www.ioc.ee/cnw09

18–21 October, New Paltz, NY, USA. **IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2009).**

26–28 October, Edinburgh, UK. **Euronoise 2009.** Web: http://www.euronoise2009.org.uk

26–30 October, San Antonio, TX, USA. **158th Meeting of the Acoustical Society of America.** Web: http://asa.aip.org/meetings.html

05–06 November, Dübendorf, Switzerland. **Swiss Acoustical Society Autumn Meeting.** Web: http://www.sga-ssa.ch

23–25 November, Adelaide, Australia. **Australian Acoustics Society National Conference.** Web: http://www.acoustics.asn.au/joomla

2010

15–19 March, Dallas, TX, USA. **International Conference on Acoustics, Speech, and Signal Processing.** Web: http://icassp2010.org

19–23 April, Baltimore, MD, USA. Joint Meeting: **158th Meeting of the Acoustical Society of America and Noise Con 2010.** Web: http://asa.aip.org/meetings.html

09–11 June, Aalborg, Denmark. **14th Conference on Low Frequency Noise and Vibration.** Web: http://lowfrequency2010.org

13–16 June, Lisbon, Portugal. **INTERNOISE2010.** Web: http://www.internoise2010.org

23–27 August, Sydney, Australia. **International Congress on Acoustics 2010.** Web: http://www.ica2010sydney.org

14–18 September, Kyoto, Japan. 5th **Animal Sonar Symposium** Web: http:cse.fra.affrc.go.jp/akamatsu/AnimalSonar.html

15–18 September, Ljubljania, Slovenia. **Alp-Adria-Acoustics Meeting joint with EAA.** E-mail: mirko.cudina@fs.uni-lj.si

26–30 September, Makuhari, Japan. **Interspeech 2010—ICSLP.** Web: http://www.interspeech2010.org

11–14 October, San Diego, California, USA. **IEEE 2010 Ultrasonics Symposium.** E-mail: bpotter@vectron.com

14–16 October, Niagra-on-the-lake, Ont., Canada. **Acoustics Week Canada**. Web: http://caa-aca.ca/E/index.html

15–19 November, Cancun, Mexico. **2nd Iberoamerican Conference on Acoustics (Joint Meeting of the Acoustical Society of America, Mexican Institute of Acoustics, and Iberoamerican Federation on Acoustics), Cancun, Mexico.** Web: http://asa.aip.org/meetings.html

2011

27 June–01 July, Aalborg, Denmark. **Forum Acusticum 2011.** Web: http://www.fa2011.org

27–31 August, Florence, Italy. **Interspeech 2011.** Web: http://www.interspeech2011.org

04–07 September, Gdansk, Poland. **International Congress on Ultrasonics.** Web: TBA

04–07 September, Osaka, Japan. **Internoise 2011.** Web: TBA

2013

02–07 June, Montréal, Canada. **21st International Congress on Acoustics (ICA 2013).** Web: http://www.ica2013montreal.org

REVIEWS OF ACOUSTICAL PATENTS

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

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

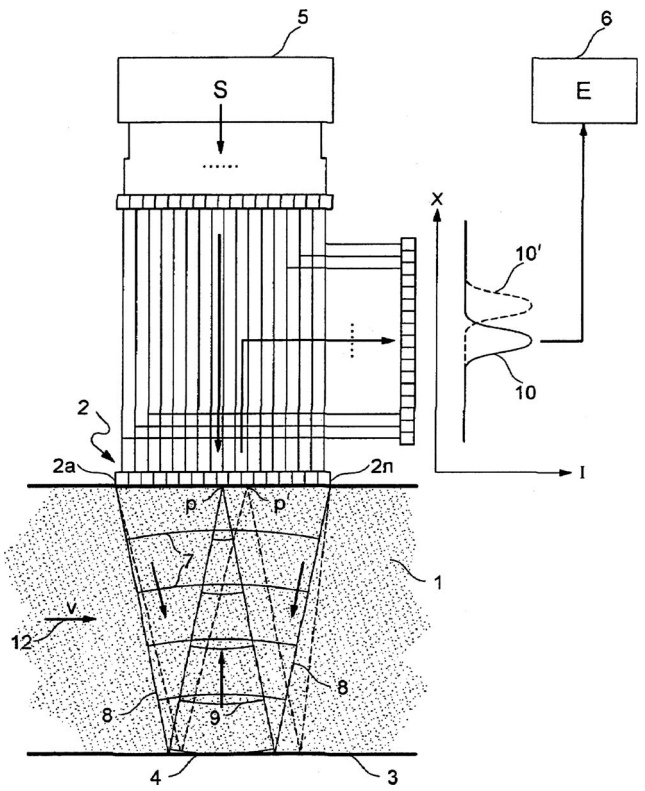
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prairie, Minnesota 55344*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
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*

7,503,225

43.28.Py ULTRASONIC FLOW SENSOR HAVING A TRANSDUCER ARRAY AND REFLECTIVE SURFACE

Tobias Lang, assignor to Robert Bosch GmbH
17 March 2009 (Class 73/861.25); filed in Germany 18 March 2004

The authors describe an ultrasonic measurement sensor utilizing the drag of sound waves by a moving medium. The transducer is composed of a number of individual transducers (2a–2n) formed in a linear array. The array is then put in contact with one side of a fluid-carrying pipe, the other side having a reflecting (smooth) surface to reflect the incident beam back at the source array. By electrically delaying elements of the source array, a converging wedge-shaped beam is formed that illuminates a small region of the array upon its return. The position of the illuminated spot is linearly dependent on the flow velocity and can be used to measure it.—JAH



7,408,841

43.30.Yj SYSTEM AND METHOD FOR CALCULATING THE DIRECTIVITY INDEX OF A PASSIVE ACOUSTIC ARRAY

John R. Welch *et al.*, assignors to The United States of the America as represented by the Secretary of the Navy
5 August 2008 (Class 367/131); filed 27 July 2007

This software patent just illustrates how to compute the directivity index for a wide variety of underwater arrays. There is nothing new here.—MK

7,411,333

43.35.Pt SURFACE ACOUSTIC WAVE DEVICE

Yoshikazu Kihara *et al.*, assignors to TDK Corporation
12 August 2008 (Class 310/313 B); filed in Japan 21 July 2005

Optimum lithium tantalate and lithium niobate crystal cut angles are claimed.—AJC

7,384,670

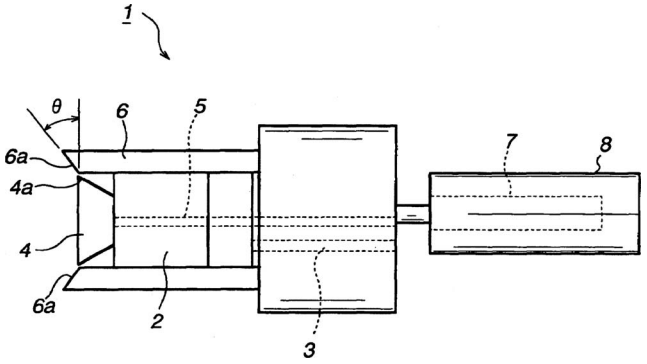
43.35.Zc COATING METHOD AND ATOMIZER

Shinji Tani *et al.*, assignors to Ransburg Industrial Finishing K.K.
10 June 2008 (Class 427/421.1); filed in Japan 27 March 2003

Automobile coating (paint) application device 1 is claimed where pipe 5 supplies paint onto rotating head 4 which throws paint droplets onto ul-

 0001-4966/2009/126(2)/927/15/$25.00

trasonic vibrator face 6a. The paint droplets are atomized on contact with



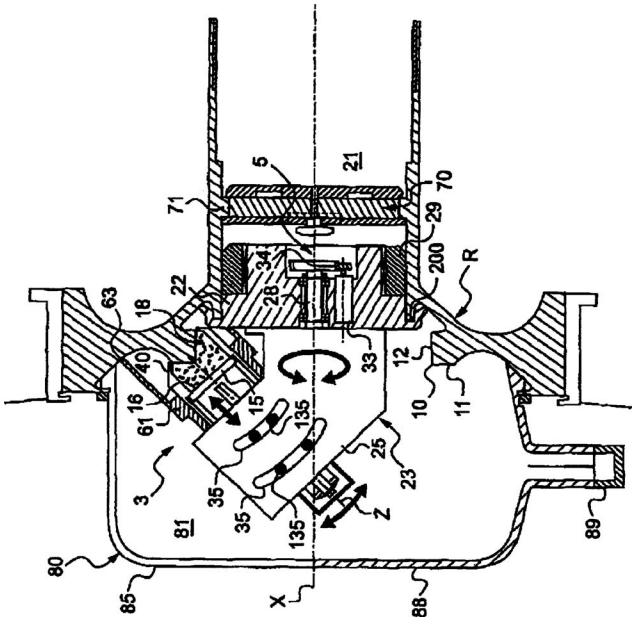
face 6a that is also slanted at angle θ. A weak current of air or a charging voltage transports atomized paint onto the surface to be coated. Banks of such sprayers 1 surround an automobile body that is passed before them in production.—AJC

7,389,663

43.35.Zc ACOUSTIC SHOT PEENING METHOD AND APPARATUS

Patrick Cheppe *et al.*, assignors to Sonats-Societe des Nouvelles Applications des Techniques de Surfaces
24 June 2008 (Class 72/53); filed in France 20 October 2006

An acoustically driven shot peening apparatus 81 is claimed where a vibrating tool 3 carrying "Sonotrode" 15 having vibrating (straight arrow) surface that propels pellets 18 to repeatedly strike surface 10 that is to be toughened. Tool 3 is traversed around surface 10 by rotating carrier 5 (round arrow) on axis X to toughen the entire circumference of 10. No sonic frequency is cited. Since the transducer style specified is piezoelectric, it might be high frequency audio or ultrasonic.—AJC

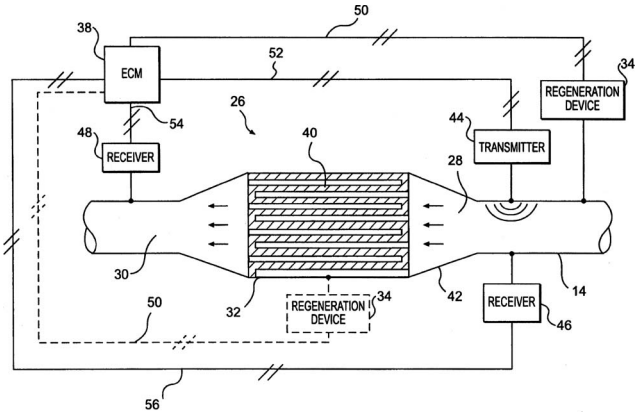


7,395,710

43.35.Zc SYSTEM AND METHOD FOR MONITORING A FILTER

Mohamed I. Daoud, assignor to Caterpillar Incorporated
8 July 2008 (Class 73/584); filed 28 February 2006

A diesel engine exhaust particulate filter condition monitor is claimed where attenuation of the sound or ultrasound transmitted by 44 into exhaust duct 28 through filter 40 and to receiver 48 is compared to the direct sound received by receiver 46. Similar sound intensity at both receivers indicates a clean filter or a filter that has cracks, holes, or leaks. A reduced sound intensity at 48 indicates a clogged filter that should be cleaned. Suitable sound frequencies include audible sound and ultrasound.—AJC



7,398,685

43.35.Zc MEASURING METHOD USING SURFACE ACOUSTIC WAVE DEVICE, AND SURFACE ACOUSTIC WAVE DEVICE AND BIOSENSOR DEVICE

Atsushi Itoh and Motoko Ichihashi, assignors to Ulvac, Incorporated
15 July 2008 (Class 73/599); filed in Japan 11 June 2004

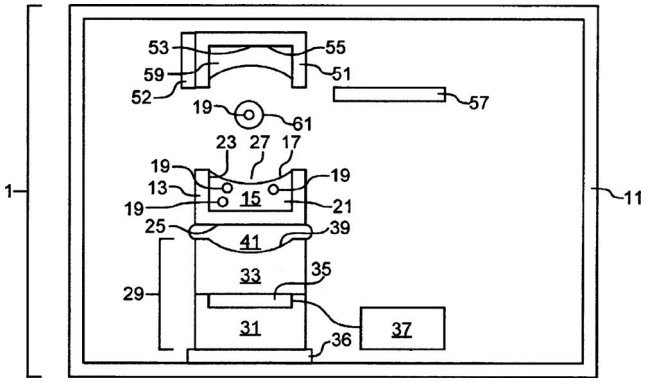
Compensating means for the effect of substance viscosity on the measurement of biochemical substance density while in contact with a surface acoustic wave (SAW) is claimed. Two SAW operating frequencies are used, one above and one below a center frequency. The frequency difference between the upper and lower frequency SAWs required to cause a phase difference of just plus and minus 45 deg with respect to the center frequency is used to calculate and compensate for the viscosity effect on aforementioned density measurement.—AJC

7,405,072

43.35.Zc ACOUSTIC RADIATION FOR EJECTING AND MONITORING PATHOGENIC FLUIDS

Mitchell W. Mutz and Richard N. Ellson, assignors to Picoliter Incorporated
29 July 2008 (Class 435/286.3); filed 18 July 2002

Ultrasonic apparatus 1 for selecting pathogen containing particle 19 and ejecting it as droplet 61 of carrying fluid for observation is claimed. Controller 37 provides electric driving current to generator 31 having transducer 35 which projects ultrasound through coupler 33 where spherical surface 41 provides focusing of that ultrasound onto fluid surface 41. Reflected



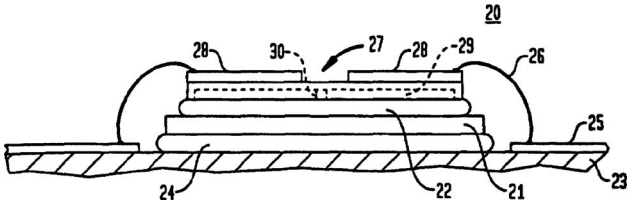
ultrasound serves to signal particle 19 presence near surface 27. Increased drive by controller 37 then causes intense ultrasound at surface 27 to eject droplet 61 carrying particle 19 to travel to well 51–55 with cover 57 removed to collect droplets 19 for analysis.—AJC

7,397,165

43.38.Fx SURFACE WAVE RESONATOR WITH REDUCED ACCELERATION SENSITIVITY

John A. Kosinski, assignor to The United States of America as represented by the Secretary of the Army
8 July 2008 (Class 310/313 R); filed 27 March 2007

Reduced acceleration sensitivity of surface acoustic wave (SAW) device 20 is claimed where stiffener plate 21 is bonded by adhesive 22 to piezoelectric substrate 27. The mass of substrate 27 is also reduced by micromachining void 29, while being stiffened by cross-beam 30 left behind in the micromachining operation.—AJC

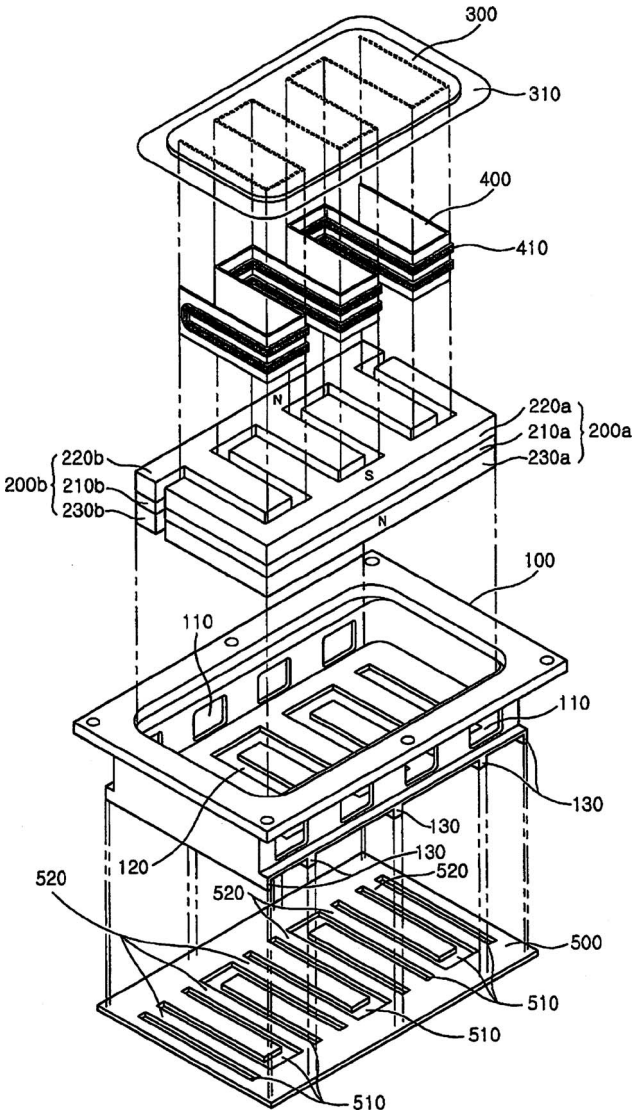


7,502,486

43.38.Ja PLANE SPEAKER HAVING COIL PLATE GUIDE DEVICE

Joung-Youl Shin, Naeson-Dong, Uiwang 437-804 and Byung-Wan Han, 1478-4 Seocho3-Dong, Seocho-Gu, Seoul 137-868, both of Republic of Korea
10 March 2009 (Class 381/396); filed in Republic of Korea 5 December 2003

A multi-turn, meandering coil planar loudspeaker effectively has two windings on a single former, each winding operating in its own magnetic gap. As a result, the voice coil assembly is quite deep and is subject to deformation over time. This variant adds a lightweight, flexible plate at the bottom that acts like a centering spider to maintain coil alignment and prevent scraping.—GLA

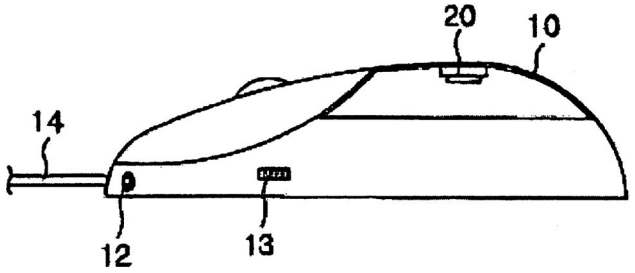


7,505,032

43.38.Ja MOUSE DEVICE HAVING VOICE OUTPUT UNIT

Jong-Hyun Shin, assignor to Soundscape Company, Limited
17 March 2009 (Class 345/163); filed in Republic of Korea 16 April 2003

Miniature loudspeaker 20 is attached to the interior of computer mouse casing 10. There is no sound exit opening. Instead, the variations in air pressure excite corresponding vibrations in the casing, allowing the sound to be both heard and felt.—GLA

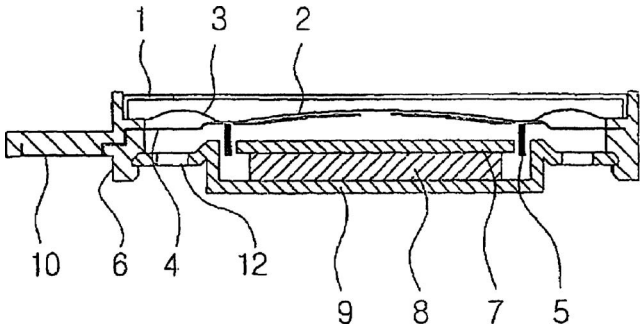


7,505,603

43.38.Ja DYNAMIC MICRO SPEAKER WITH DUAL SUSPENSION

Ok-Jung Yoo, assignor to Jin Young Acoustic Company, Limited
17 March 2009 (Class 381/404); filed in Republic of Korea 30 August 2002

This miniature loudspeaker employs a dual suspension to maintain coil alignment during long excursions. Upper suspension 3 is conventionally formed as part of the diaphragm. Lower suspension 4 is highly resilient. If made of electrically conductive material it can also provide two electrical paths from the ends of the moving coil to the stationary connection terminals.—GLA

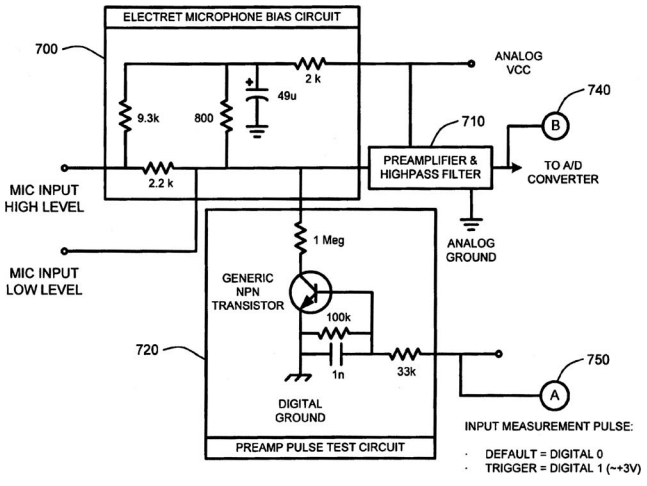


7,428,309

43.38.Kb ANALOG PREAMPLIFIER MEASUREMENT FOR A MICROPHONE ARRAY

Henrique Malvar and Ivan Tashev, assignors to Microsoft Corporation
23 September 2008 (Class 381/58); filed 4 February 2004

From very digital Microsoft comes this very analog disclosure in order to calibrate the preamplifiers and filtering in a microphone array and add an additional signal path as shown in the figure. This can be used as the input of a calibrated pulse. Treated as an impulse, this can be used as the basis for a spectral analysis method such as the fast Fourier transform, and then an inverse compensation filter can be calculated.—MK

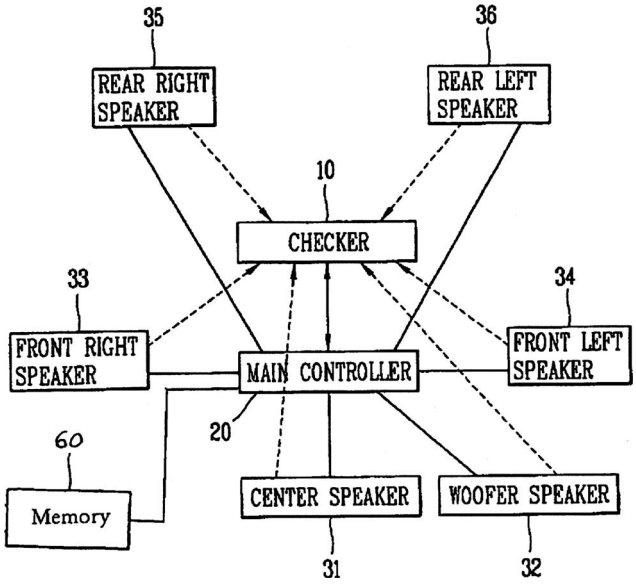


7,428,310

43.38.Kb AUDIO OUTPUT ADJUSTING DEVICE OF HOME THEATER SYSTEM AND METHOD THEREOF

Jeong-Hyun Park, assignor to LG Electronics Incorporated
23 September 2008 (Class 381/104); filed in Republic of Korea 31 December 2002

Imagine: Add a microphone to a multichannel audio system and you can record the levels and the delays at a listening spot. Accordingly, compensation could be computed, but then you would have to give an algorithm to do so, which is missing from this Lucky Goldstar (LG Electronics) patent. Also missing are all the possible different implementations. All in all, without these details, this borders on the obvious.—MK

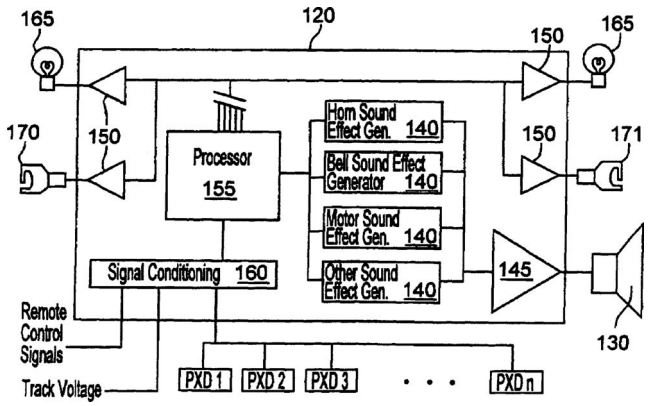


7,429,931

43.38.Md PROXIMITY CONTROL OF ON-BOARD PROCESSOR-BASED MODEL TRAIN SOUND AND CONTROL SYSTEM

Frederick E. Severson, Beaverton, Oregon
30 September 2008 (Class 340/686.1); filed 30 June 2005

In the quest for more realistic sound generation in model railroading, the inventor uses the now ubiquitous microprocessor with various sound generators as shown. Notice how the track voltage can be used as a control input and thereby generate speed specific sounds such as brake squeals and motor sounds.—MK

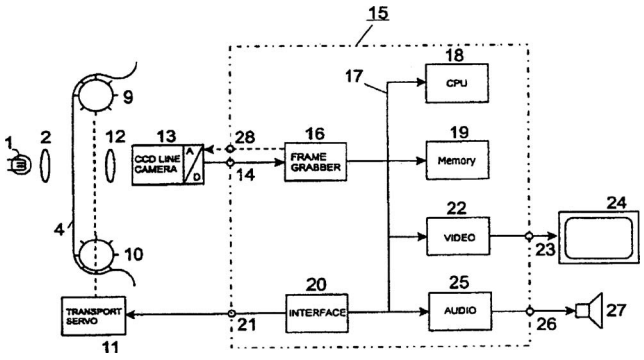


7,400,821

43.38.Ne DEVICE AND METHOD FOR REPRODUCING PHOTOGRAPHIC SOUND RECORDINGS

Dieter Poetsch *et al.*, assignors to Thomson Licensing
15 July 2008 (Class 386/104); filed in Germany 11 September 2000

35 mm film sound can be recorded either optically or magnetically. In this case, the disclosure describes an image processing algorithm that digitizes the optical sound track and finds the edges so that the audio signal magnitude can be computed. Notably, the signal processing equations are provided.—MK

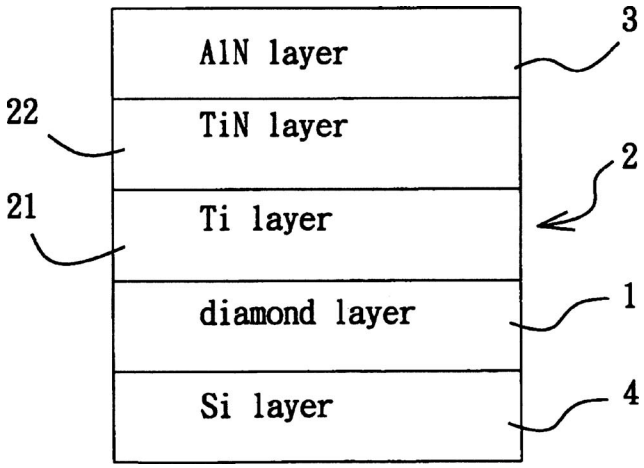


7,375,453

43.38.Rh SURFACE ACOUSTIC WAVE SUBSTRATE

I-Nan Lin *et al.*, assignors to Tamkang University
20 May 2008 (Class 310/313 R); filed in Taiwan 3 April 2006

High propagation speed surface acoustic wave devices can be made from a crystalline aluminum nitride (AlN) layer 3 on a diamond layer 1, but delamination can occur for AlN layers greater than 2 μm thick. The authors claim that graded adhesion layers of titanium 21 and titanium nitride 22 will enhance the adhesion of AlN to diamond.—AJC

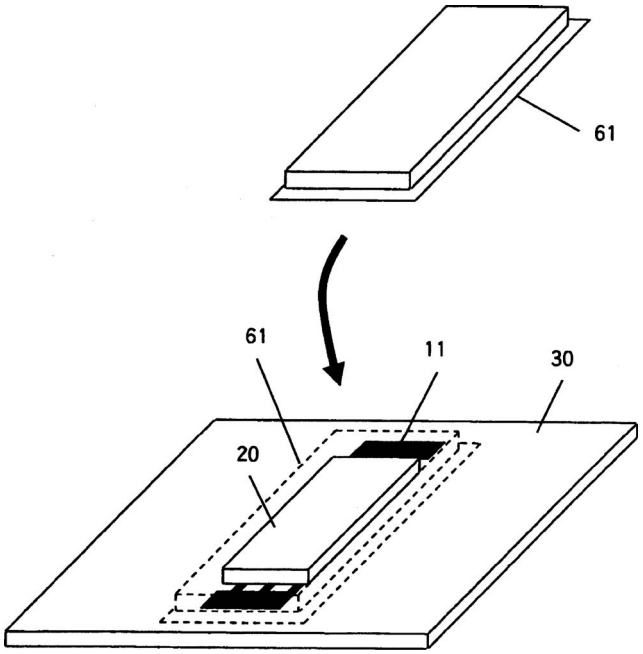


7,375,454

43.38.Rh SURFACE ACOUSTIC WAVE EXCITATION DEVICE

Masaya Takasaki, assignor to National University Corporation Saitama University
20 May 2008 (Class 310/313 R); filed in Japan 19 October 2004

The author claims a surface acoustic wave device 11-20 that also acts as a vibration transducer to launch a surface wave into glass substrate 30 that is also a touch-sensitive screen. Transducer cover 61 is held to glass panel 30 by a vacuum that is created during manufacture.—AJC

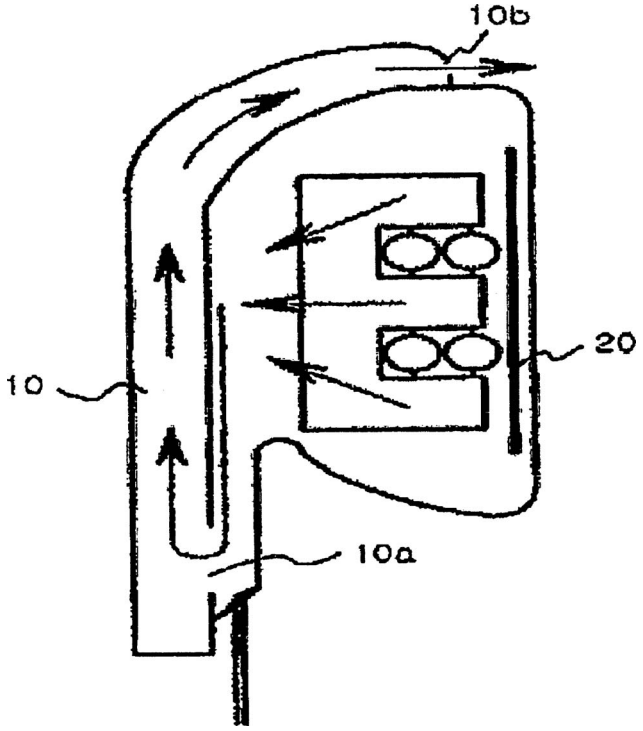


7,499,562

43.38.Si VIRTUAL MULTI-CHANNEL SPEAKER UNIT

Soon-Teak Shim, Daejeon City, Republic of Korea
3 March 2009 (Class 381/382); filed in Republic of Korea 22 September 2003

Many patents are difficult to understand, but in this case both the patent document and the invention itself are vague and ambiguous. The text and illustrations describe a "speaker unit" that is actually an earphone designed as a miniature vented loudspeaker system. The vent takes the form of a long tube 10, "…physically shaped into a specific form to form a virtual



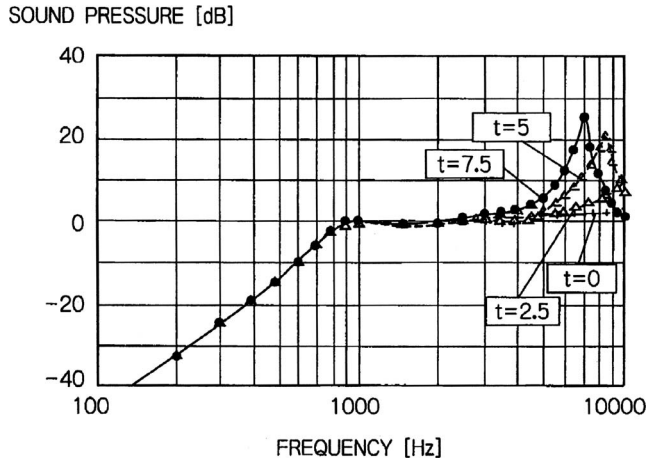
channel of the speaker and which can cancel air pressure applied to a user's eardrum to reduce fatigue and improve sound quality of the speaker." It appears that "multi-channel" refers to the front and rear sound channels rather than multi-channel program material, yet the patent also makes reference to creating a "multi-channel effect." The claims define "a virtual multi-channel speaker unit," with no mention of ears, eardrums, or earphones. However, the speaker unit does contain a "solenoid coil," "a magnetic body around which the solenoid coil is wound," and "a diaphragm that faces one side of the magnetic body."—GLA

7,499,735

43.38.Si MOBILE STATION WITH SOUND SPEAKER

Yukio Murata, assignor to NEC Corporation
3 March 2009 (Class 455/575.1); filed in Japan 30 March 2004

The goal of this patent is to simplify the construction of a cellular telephone by incorporating a sound-conducting tube for the loudspeaker as part of the inner frame. Everything simply snaps into place. The inventor's



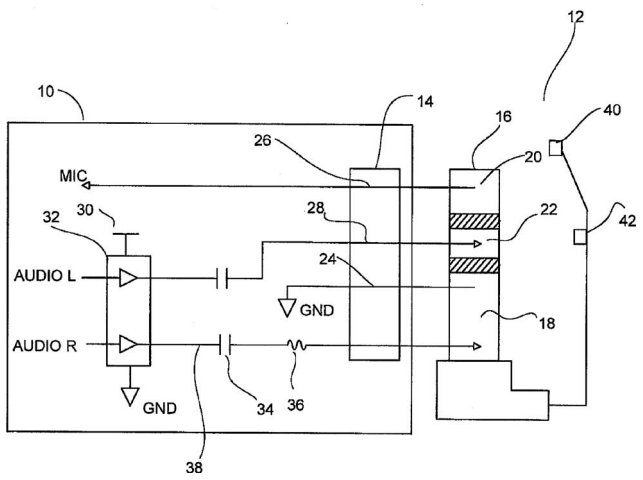
curiosity led him to measure the response of the assembly with tubes of different lengths. Although not mentioned in the patent claims, it is apparent that tube geometry can be used to shape the high frequency response of the system.—GLA

7,502,478

43.38.Si STEREO AUDIO HEADSET INTERFACE AND METHOD THAT IS COMPATIBLE WITH MONO HEADSETS

Jay King, assignor to Kyocera Wireless Corporation
10 March 2009 (Class 381/55); filed 21 November 2003

Test equipment and line-level audio amplifiers often include an output series resistor to provide short circuit protection. If the device is a stereo headphone amplifier then the two series resistors will also allow a mono headset to be used. This seemingly obvious observation so enchanted the patent examiner that 11 claims were granted.—GLA

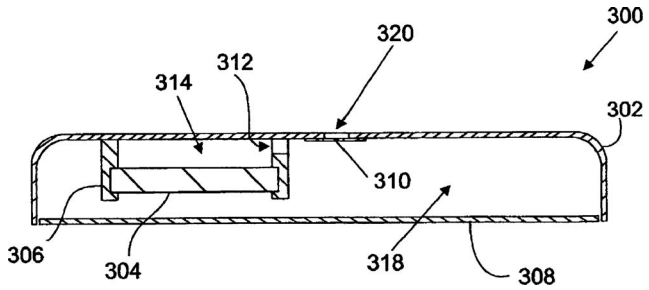


7,505,602

43.38.Si MOBILE DEVICE WITH IMPROVED ACOUSTIC PORTING

W. Chris Eaton, assignor to Sony Ericsson Mobile Communications AB
17 March 2009 (Class 381/351); filed 28 February 2003

In this mobile telephone, loudspeaker 304 drives front chamber 314 which is vented through port 312. High frequency response can be contoured by adjusting the combination of chamber volume and vent geometry. The interior cavity 318 is used as a "mixing volume" to combine front and rear radiations. Sound exits through opening 320. The patent includes a response curve showing rolloff above 1 kHz, but peaking again at about 3.2 kHz. Usable response extends from roughly 150 Hz to about 3.5 kHz.—GLA

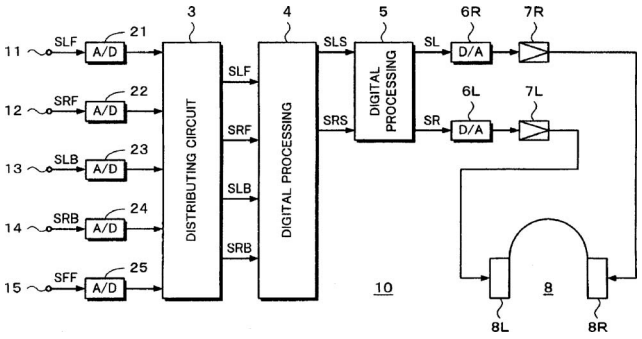


7,502,477

43.38.Vk AUDIO REPRODUCING APPARATUS

Kiyofumi Inanaga and Yuji Yamada, assignors to Sony Corporation
10 March 2009 (Class 381/17); filed in Japan 30 March 1998

Exactly what is patented here, why it is patented, and what is to be accomplished by the invention are all questions that remain unanswered after a third reading of the patent document. Inputs 11–15 represent five



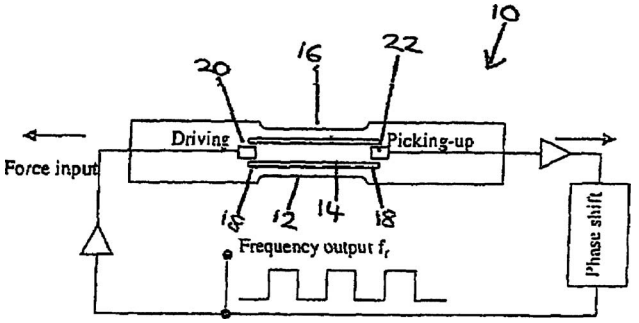
channels of a surround sound system, except that the front-center channel seems to be missing. Claim 1 adds to the confusion by listing the previous five channels and then adding, "…an additional channel signal being the center channel." Going back to the diagram, distributing circuit 3 presumably combines the low-frequency effect signal with the four main signals. Digital processing unit 4 then utilizes panning functions to reduce the four signals to two for headphone listening. That is how I interpret the patent but I may be wrong.—GLA

7,498,721

43.40.Cw RESONANT SENSOR ASSEMBLY

Edward Barry Jones *et al.*, assignors to Brunel University
3 March 2009 (Class 310/338); filed in United Kingdom 5 February 2003

This patent describes an interesting form of the tuning fork resonator that is designed to be simple to fabricate and mount to a fixture. The basic resonator is three parallel beams 12, 14, and 16 that are used in a mode of vibration wherein 12 and 16 move out of phase with central beam 14. Beam 14 is twice the width of beams 12 and 16, a ratio that the authors say maximizes the Q by minimizing coupling of vibrations to the clamped support areas on the ends. The motion of the beams is excited by piezoelectric elements that are screen printed on areas 20 and 22 where the central beam bending stresses are highest, transducer 20 being used as a driver and 22 being used as a receiver. Using electrical feedback between the two elements, it is possible to create an oscillator with a Q of 3300 at 6 kHz in air, using stainless steel of length × width dimensions of 15×4 mm^2.—JAH

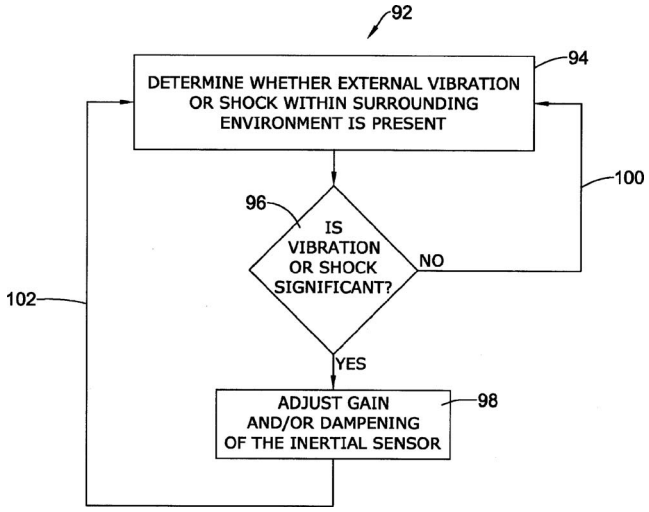


7,401,515

43.40.Jc ADAPTIVE CIRCUITS AND METHODS FOR REDUCING VIBRATION OR SHOCK INDUCED ERRORS IN INERTIAL SENSORS

Mark W. Weber, assignor to Honeywell International Incorporated
22 July 2008 (Class 73/504.12); filed 28 March 2006

Adaptive control 92 against the disturbance effect of shock and vibration on an inertial navigation sensor (not shown) is claimed, where the shock is determined by accelerometer 94. The inertial sensor (not shown) output 102 is monitored by 96 for out-of-likelihood deviations 100. On detecting an unacceptable deviation (YES), the gain and/or the damping 98 of sensor output signal is adjusted to minimize that error for guidance purposes.—AJC

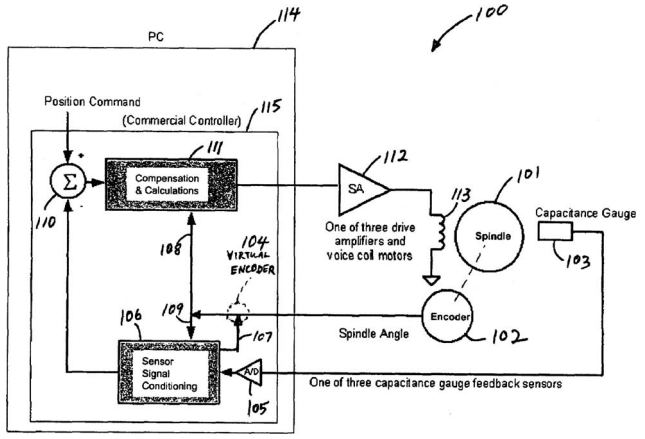


7,370,524

43.40.Vn ADAPTIVE VIBRATION CONTROL USING SYNCHRONOUS DEMODULATION WITH MACHINE TOOL CONTROLLER MOTOR COMMUTATION

David James Hopkins, assignor to Lawrence Livermore National Security, LLC
13 May 2008 (Class 73/146.2); filed 15 August 2005

An active vibration control system for reducing the housing vibration due to imbalance of an internal rotating spindle is claimed. The signal produced by capacitance displacement sensor 103 is fed into processor 115 along with the phase signal from encoder 102 into synchronous demodulator 106 to produce position signal 110. Comparator 110 provides an error signal to amplifier 112 that drives transducer coil 113 to reduce spindle vibration.—AJC



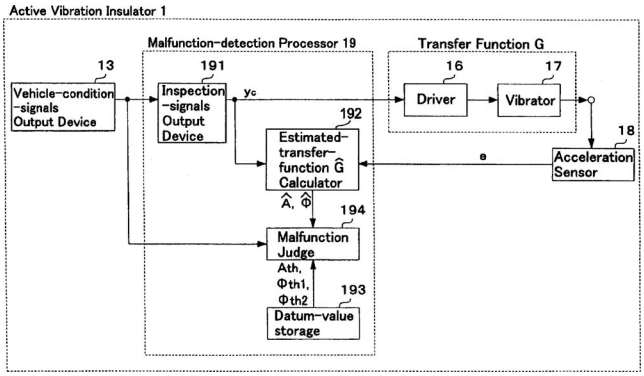
7,403,367

43.40.Vn ACTIVE VIBRATION INSULATOR

Takayoshi Yasuda *et al.*, assignors to Tokai Rubber Industries, Limited
22 July 2008 (Class 361/139); filed in Japan 28 December 2004

A vehicle engine active vibration isolation control system 1 is claimed that will inspect (191–194) and identify a vibrator 16-17 malfunction drive

signal yc characteristic of a broken wire, locked plunger, ruptured force plate, etc. The inspection 191 and adjustment rates are preferably below 25 per second so as to be inaudible to vehicle passengers. The inspection period when examining transducer drive currents is less than 1 s and preferably between 0.8 and 0.4 s.—AJC

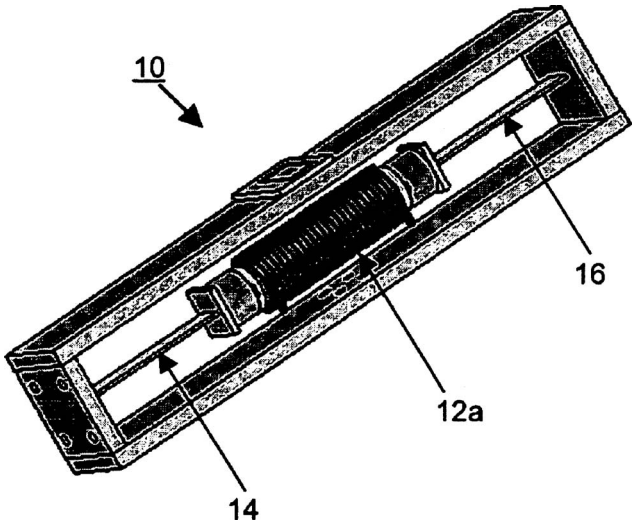


7,410,039

43.40.Vn TUNABLE VIBRATION ABSORPTION DEVICE

Siu Wing Or *et al.*, assignors to ASM Assembly Automation Limited
12 August 2008 (Class 188/380); filed 4 February 2005

Smart vibration dynamic damper system 10 is claimed where vibration transverse to support beam 14–16 sets mass 12c on tensioned beam 14–16. Significant vibration energy is absorbed when the beam tension is appropriate for mass-spring resonance. Beam tension is adjusted by a variable current through the spiral coil around mass 12a, causing the magnetostrictive central material part of beam 14–16 to shrink in length. A piezoelectric tension sensor (not shown) is incorporated into the beam assembly inside the coil. Current is adjusted by a controller to cause maximum mass-spring vibration amplitude for maximum vibration damping.—AJC



7,482,731

43.40.Vn KIT AND METHOD FOR CONSTRUCTING VIBRATION SUPPRESSION AND/OR SENSING UNITS

Baruch Pletner and Gregory A. Zvonar, assignors to IPTrade, Incorporated
27 January 2009 (Class 310/321); filed 27 October 2005

The kit described in this patent pertains to an active vibration suppression system that is intended to be mounted at key locations, such as the effector of a robotic arm, of equipment whose performance is highly sensitive to vibration. The system is designed in modular fashion and requires no cables (or only small cables) to be installed between the equipment's stationary and moving parts. The system consists of one or more sensing devices, a controller, and one or more actuators. Each of these is a self-contained unit that includes all electronics on printed circuit boards and that communicates with the others wirelessly as far as possible. Suitable control algorithms supposedly can be programmed into the circuit boards but are not discussed in this patent.—EEU

7,497,146

43.40.Vn METHOD FOR FILTERING ROTATIONAL VIBRATIONS FOR MOTOR VEHICLE

Jacques Clausin, assignor to Valeo Embrayages
3 March 2009 (Class 74/574.1); filed in France 4 May 2001

This rotational dynamic absorber is intended to suppress shaft vibrations that occur in a limited range of frequencies. It is constructed so that it has its natural frequency in the range of frequencies of concern and is provided with an active control system that tunes it to the particular frequency at which the shaft vibrates at a given time. This tuning is done via inductive windings located near the absorber's periphery. These windings interact with permanent magnets attached to the absorber edge. The current in the inductive windings is determined by a controller, which acts on the basis of the output of a shaft-mounted vibration sensor.—EEU

7,478,563

43.40.Yq APPARATUS AND METHOD TO MEASURE CABLE TENSION

Simon Weisman, Toronto, Ontario, Canada
20 January 2009 (Class 73/831); filed 23 June 2006

The tension in a cable is determined by measuring the cable's natural fundamental frequency. The system described in this patent consists of a pair of clamps that have an accelerometer attached and that may be easily affixed to cables of various diameters. The accelerometer signal is fed to a computer, which determines the signal's frequency and calculates the cable tension by means of a simple formula. This formula includes the length and the weight of the cable, which presumably must be supplied by the user of the apparatus. The patent does not indicate how the cable is set into vibratory motion.—EEU

7,487,679

43.40.Yq VEHICLE VIBRATION ANALYZER

Scott Anthony Sirrine and Kevin W. Marsh, assignors to Eaton Corporation
10 February 2009 (Class 73/660); filed 10 May 2005

This analyzer system is intended to permit a non-specialist to identify the sources of excessive vibrations in a vehicle. It makes use of signals from a number of vibration sensors. These signals are fed to a computer that attributes the vibrations to common sources, compares the vibrations to predetermined threshold values, and causes possible sources of excessive vibrations and corresponding repair recommendations to be displayed on a screen. The patent includes signal analysis flowcharts, but no further information on data processing.—EEU

7,493,238

43.40.Yq ABNORMAL NOISE CORRECTION VERIFICATION APPARATUS

Tsutomu Tanaka *et al.*, assignors to Nissan Motor Company, Limited
17 February 2009 (Class 702/191); filed in Japan 30 January 2006

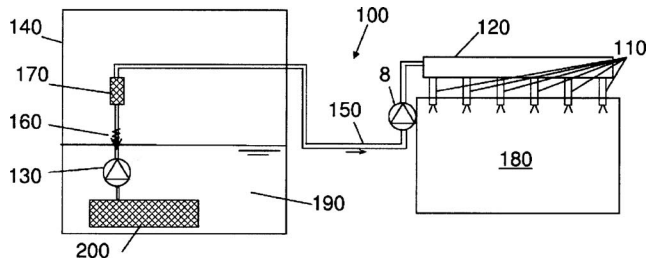
The apparatus described here is intended to verify whether action taken to correct an abnormal noise in a motor vehicle is effective. The signals obtained from one or more vibration sensors before and after the corrective action under identical operating conditions are stored, filtered, and compared.—EEU

7,373,924

43.50.Gf METHOD AND SYSTEM TO MITIGATE PUMP NOISE IN A DIRECT INJECTION, SPARK IGNITION ENGINE

Eric Krengel *et al.*, assignors to Ford Global Technologies, LLC
20 May 2008 (Class 123/457); filed 10 May 2007

Objectionable ticking noise from engine 180 fuel injectors 110 emitted at low speed and idle is reduced by disabling the high pressure action of pump 8. Low fuel pressure from tank pump 130 is sufficient for low engine power conditions, whereby the fuel injection metering can also be more accurate.—AJC

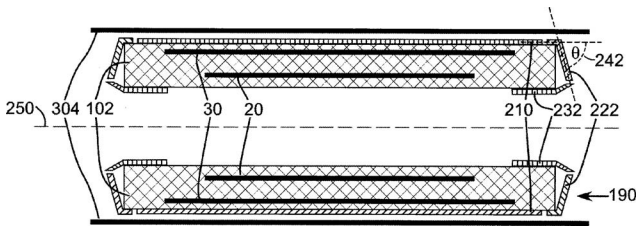


7,375,526

43.50.Gf ACTIVE-PASSIVE ELECTROMAGNETIC SHIELDING TO REDUCE MRI ACOUSTIC NOISE

William A. Edelstein *et al.*, Schenectady, New York
20 May 2008 (Class 324/318); filed 20 October 2006

Improved noise reduction for magnetic resonance imaging (MRI) devices is claimed. Metal enclosure 304 is located around the entire apparatus. MRI devices work by superimposing an axial radio frequency field, scan-modulated at audio frequencies, via "gradient" coils 20 and 30. A metal tube, not shown, large enough in diameter to pass a human body, is located along axis 250. When the gradient rf field is not strictly aligned along axis 250, the rf field component normal to the metal surface of the tubes will induce rf eddy currents, also modulated at an audio frequency. In conjunction with the strong static magnetic field necessary to produce magnetic resonance, this eddy current results in a Lorentz force on metal surfaces 304 and the axial central tube, causing these metal surfaces to vibrate at audio frequencies and to emit strong acoustic noises. Gradient coil 30 is extended axially to reduce the radial field component or fringing and is named the "active" portion of noise reduction. Passive reduction in such fringing and noise was achieved in prior art by adding a conducting metal (viz., copper) tube just outside gradient coil 30. The claimed additional passive noise reduction is achieved by adding shields 210, 222, and 232. Angle 242 is chosen to improve noise reduction.—AJC

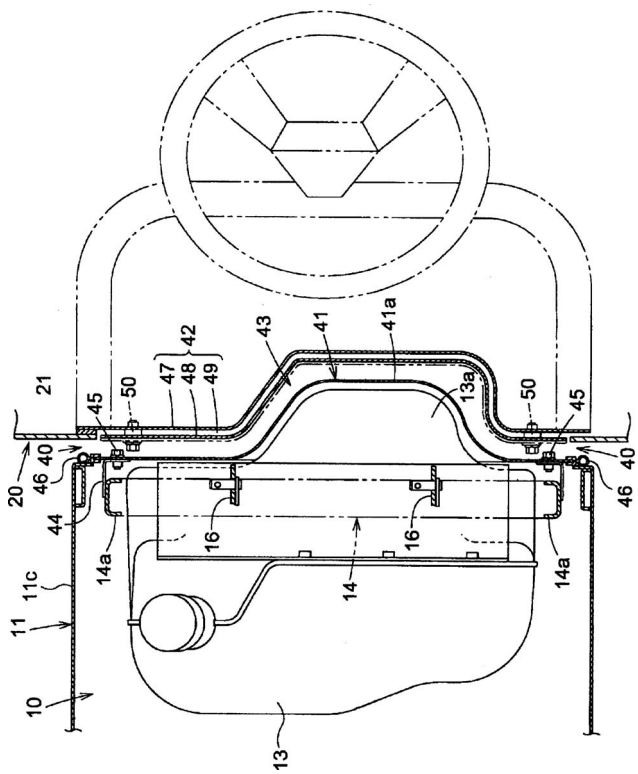


7,399,020

43.50.Gf SOUND INSULATION STRUCTURE FOR TRACTOR

Kenichi Aoyama *et al.*, assignors to Kubota Corporation
15 July 2008 (Class 296/39.3); filed in Japan 29 September 2006

Improved insulation of tractor engine noise from engine (not shown) arriving at driver is claimed where isolation gap 40 between hood 13 and cabin wall 21 is provided.—AJC



7,503,429

43.55.Ev ACOUSTIC PANEL ASSEMBLY

Dean Arden Boyce, assignor to Deere & Company
17 March 2009 (Class 181/290); filed 29 March 2007

The assembly has two parts: an outer hard sound barrier shell of polypropylene and a sound absorbing panel insert. The two panels are designed to interlock and nest, thus simplifying the installation process, reducing costs, and assuring a matched and snug fit for the inset panel. The assembly is well suited for the interior of an agricultural vehicle.—CJR

7,503,430

43.55.Ev REDUCED DUST ACOUSTIC PANEL

Mark H. Englert and Salvatore C. Immordino, Jr., assignors to USG Interiors, Incorporated
17 March 2009 (Class 181/294); filed 7 September 2006

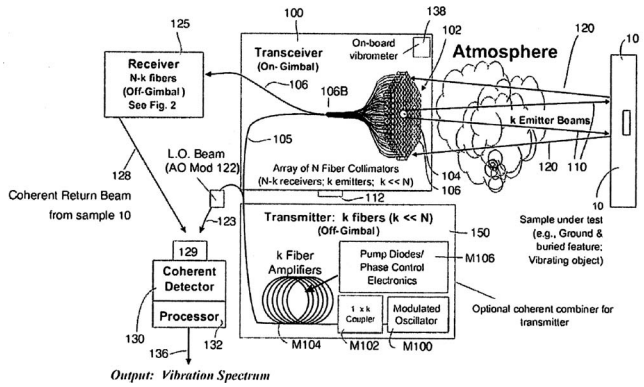
This acoustic panel has all the ingredients for sound absorption (porosity, interior voids, interlocking matrix, light weight, etc.) plus an anti-dusting additive (like polyethylene glycol) to reduce dust when the panel is cut.—CJR

7,382,465

43.60.Rw OPTICAL VIBROMETER

David M. Pepper, assignor to HRL Laboratories, LLC
3 June 2008 (Class 356/485); filed 25 August 2006

A remote sensing surface vibration monitor (figure) is claimed that transmits 1.5 μm laser radiation 110 onto surface 10. Motion or vibration by 10 will modulate the phase of reflection or back scatter radiation 120, also



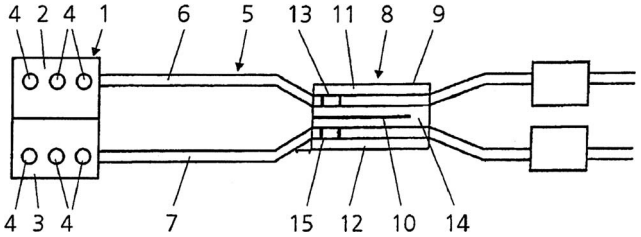
with a speckle character. Reflection or back-scattering 120 speckle is received by the fiber collimator that sends the collected radiation to receiver 125 which converts it to electric signal 128. Coherent detector 129 produces an electric signal that replicates the motion or vibration of surface 10.—AJC

7,377,359

43.66.Lj DEVICE FOR MODULATING NOISE IN A MOTOR VEHICLE

Marcus Hofmann *et al.*, assignors to Daimler AG
27 May 2008 (Class 181/275); filed in Germany 12 July 2003

Automobile V-6 and V-8 engine exhaust systems often employ a two outlet exhaust system 5 with connecting crossover 8 to efficiently provide noise reduction. However, the crossover attenuates the 1.5th order (third



harmonic of the rotation) and the 4.5th order tones necessary to produce a "sporty" (sic) exhaust sound. Maintenance of these ordered tones is achieved by adding tone resonance components 11–15 inside crossover enclosure 9.—AJC

7,495,034

43.66.Ts DEFORMABLE SOFT MOLDING COMPOSITIONS

Alan Edward Litke *et al.*, assignors to Henkel Corporation
24 February 2009 (Class 522/96); filed 20 August 2004

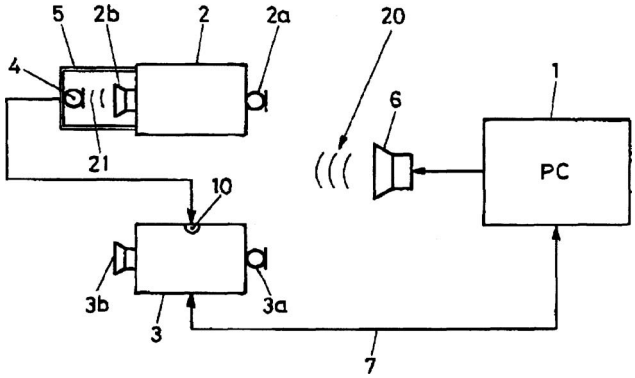
A curable composition of at least one urethane acrylate oligomer and at least one reactive diluent is used to form a deformable but durable hearing aid housing for increased wearing comfort. The hearing aid housing, which has a Shore A hardness ranging from 60 to 75, may be adhered to a soft tip by crosslinkage. Tear strength and tensile modulus achieved are said to be greater than 75 pli and 2150 psi, respectively.—DAP

7,496,205

43.66.Ts METHOD FOR ADJUSTING A HEARING DEVICE AS WELL AS AN APPARATUS TO PERFORM THE METHOD

Volker Kühnel, assignor to Phonak AG
24 February 2009 (Class 381/60); filed 9 December 2003

The goal is a more effective starting point for the hearing aid dispenser to fit a new hearing aid on a previous hearing aid wearer. This is accomplished by initially using the parameter settings of the older hearing aid(s)



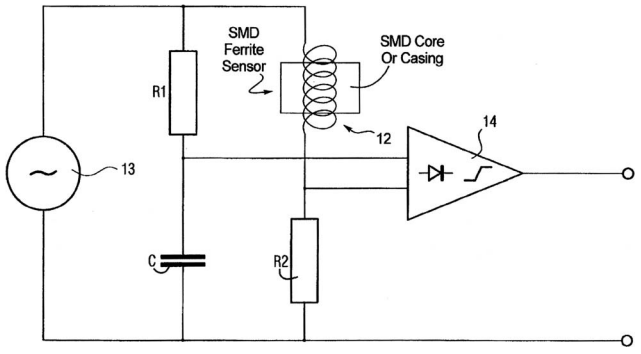
with no a priori knowledge of them. The transfer function of the older hearing device is determined with a test signal generated by the hearing aid computerized fitting-programming system. Then the processed output of the old device from a coupler microphone is used as a direct electrical input test signal to the new device so it can be adjusted to the same performance.—DAP

7,496,206

43.66.Ts HEARING AID WITH A MAGNETIC FIELD-CONTROLLED SWITCH, AND OPERATING METHOD THEREFORE

Kunibert Husung, assignor to Siemens Audiologische Technik GmbH
24 February 2009 (Class 381/312); filed in Germany 24 September 2003

A ferrite magnetic field sensor in a Wheatstone bridge circuit drives a switching circuit to alternate between modes, e.g., microphone and telecoil. The ferrite may consist of a ferrite core and coil, the latter of which is used in an LC oscillator connected to the bridge. When a static magnetic field comes into close proximity with the hearing aid, the resulting change in inductance L is sensed within the bridge circuit which triggers mode switching.—DAP

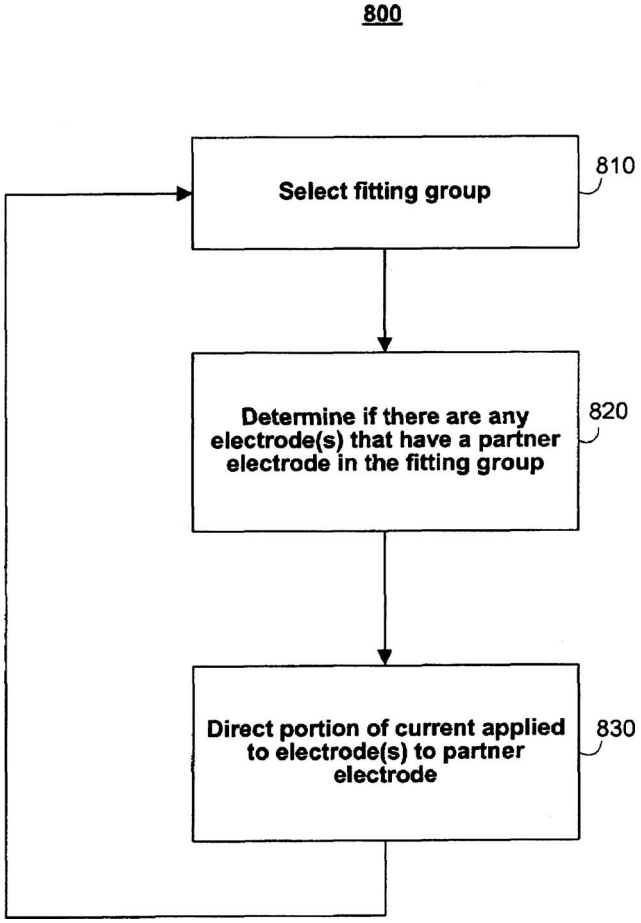


7,496,405

43.66.Ts SOUND PROCESSING AND STIMULATION SYSTEMS AND METHODS FOR USE WITH COCHLEAR IMPLANT DEVICES

Leonid M. Litvak *et al.*, assignors to Advanced Bionics, LLC
24 February 2009 (Class 607/56); filed 14 March 2005

Cochlear implant electrodes, properly stimulated, simultaneously act as virtual electrodes to more accurately reproduce spectral peak frequency information. A stimulus is applied to a first electrode to determine if a partner electrode in a fitting group associated with frequencies in a processing sub-channel would be stimulated simultaneously. All electrodes causing



this simultaneous stimulation that are not part of the fitting group are disabled and the portion of the current that would otherwise be applied to these electrodes is applied instead to their partner electrodes in the fitting group.—DAP

7,496,406

43.66.Ts SYSTEM AND METHOD FOR FITTING A COCHLEAR IMPLANT SOUND PROCESSOR USING ALTERNATIVE SIGNALS

Philip A. Segel *et al.*, assignors to Advanced Bionics, LLC
24 February 2009 (Class 607/57); filed 29 August 2003

Pulse trains using selective amounts of amplitude modulation, rather than constant amplitude stimuli, are used to set the parameters for cochlear implant stimulation in order to mimic the temporally varying signals encountered in everyday life. The temporally modulating signals may consist of noise bands with adjustable bandwidth and center frequency, complex tonal stimuli with adjustable amplitude and spectra, and actual speech tokens.—DAP

7,499,559

43.66.Ts HEARING DEVICE AND METHOD FOR CHOOSING A PROGRAM IN A MULTI PROGRAM HEARING DEVICE

Sigi Wyrsch, assignor to Bernafon AG
3 March 2009 (Class 381/314); filed in Denmark 18 December 2002

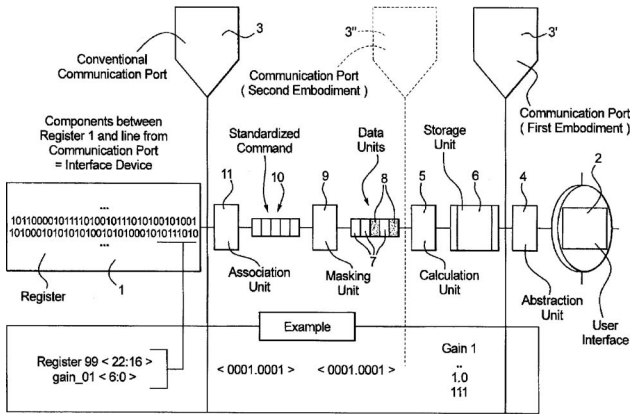
In a manual learning operating mode, signal processing parameter sets are chosen initially by the hearing device wearer for particular acoustic listening environments and are stored in the device memory along with characteristics of the listening environment. Subsequently in an automatic operating mode, the hearing device identifies the current acoustic environment as one of those previously stored and automatically chooses from memory the signal processing parameters most frequently selected by the wearer for that listening environment.—DAP

7,499,560

43.66.Ts HEARING DEVICE WITH INDIVIDUALLY CONFIGURABLE HARDWARE INTERFACE

Steffan Aschoff *et al.*, assignors to Siemens Audiologische Technik GmbH
3 March 2009 (Class 381/323); filed in Germany 27 August 2003

The goal is to reduce software overhead by making the hearing device hardware configurable. A register attached to a communication port is made available, into which an external device can store data that configures the hardware and signal processing.—DAP

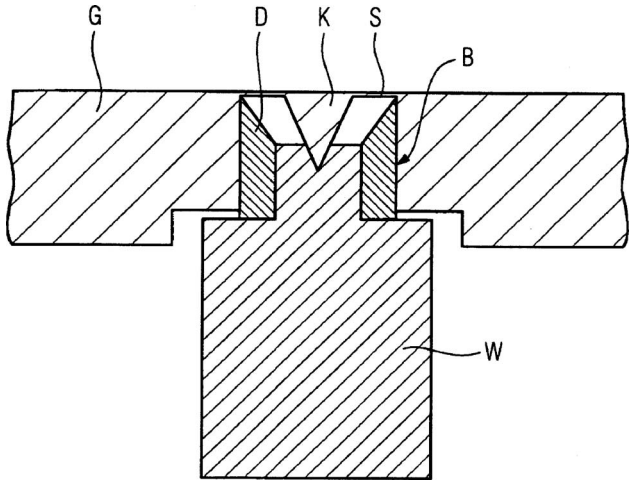


7,499,561

43.66.Ts HEARING AID WITH CERUMEN PROTECTION

Thomas Hanses and Torsten Niederdränk, assignors to Siemens Audiologische Technik GmbH
3 March 2009 (Class 381/325); filed in Germany 11 May 2004

The hearing aid output sound port is cone-shaped to facilitate easy removal of debris. The hearing aid housing and cone are covered with a hydrophobic or oleophobic protective nano-coating that has antibacterial properties owing to inclusion of silver ions.—DAP

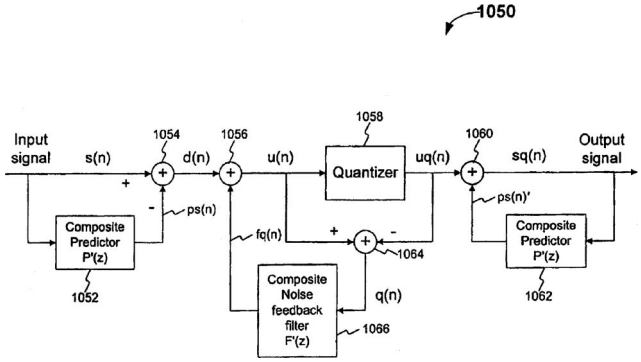


7,496,506

43.72.Gy METHOD AND APPARATUS FOR ONE-STAGE AND TWO-STAGE NOISE FEEDBACK CODING OF SPEECH AND AUDIO SIGNALS

Juin-Hwey Chen, assignor to Broadcom Corporation
24 February 2009 (Class 704/226); filed 29 January 2007

To more efficiently derive appropriate filters for use in a noise feedback coding codec structure, methodology is provided for achieving two-stage prediction and two-stage noise spectral shaping at the same time in order to achieve two-stage noise feedback coding. Two predictors are combined into a single composite predictor, or two single-stage codec structures are used in a nested manner to decouple long-term prediction and spectral shaping from short-term prediction and spectral shaping.—DAP

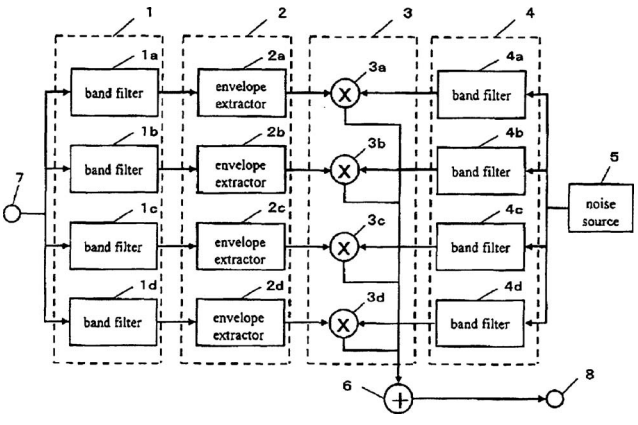


7,502,483

43.66.Ts HEARING AID, TRAINING DEVICE, GAME DEVICE, AND AUDIO OUTPUT DEVICE

Hiroshi Rikimaru, Sakyo-ku, Kyoto-shi, Kyoto, Japan
10 March 2009 (Class 381/320); filed in Japan 10 October 2002

For treating or training persons with a neural disorder, a noise-vocoded speech sound signal is generated by (1)extracting the amplitude envelopes of an incoming sound signal in multiple bands, (2) multiplying the envelopes with a noise signal extracted for the same frequency bands, and (3) summing the multiplier output signals. The patent asserts that such persons will recognize this signal by utilizing normal portions of the brain maximally.—DAP

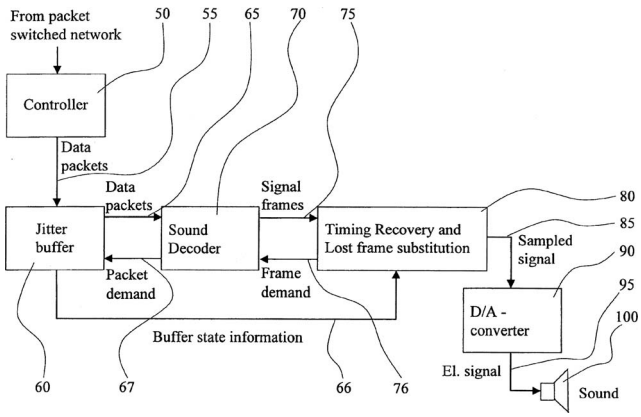


7,502,733

43.72.Gy METHOD AND ARRANGEMENT IN A COMMUNICATION SYSTEM

Soren V. Andrsen *et al.*, assignors to Global IP Solutions, Incorporated
10 March 2009 (Class 704/211); filed in Sweden 28 December 1999

This patent addresses the delay jitter problem for digitized sound packets arriving too late for continual playback of consecutive signal frames. The traditional solution to delay jitter, a jitter buffer, is not perfect because it has to be long enough to ensure a continuous packet flow while not causing too great a delay. Instead, the lengths of received signal frames are manipulated by performing time expansion of one or more signal frames with time varying lengths at time varying intervals when the jitter buffer is close to underflow.—DAP

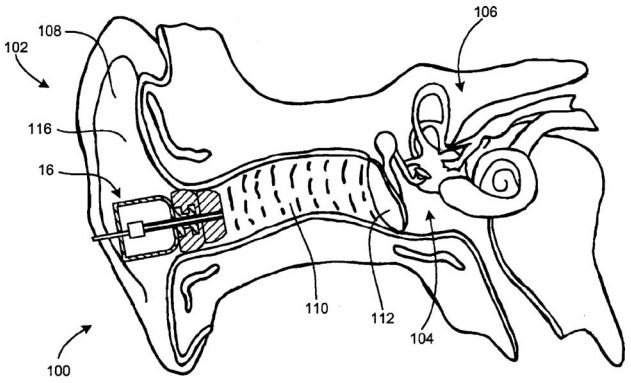


7,502,735

43.72.Gy SPEECH SIGNAL TRANSMISSION APPARATUS AND METHOD THAT MULTIPLEX AND PACKETIZE CODED INFORMATION

Hiroyuki Ehara, assignor to Panasonic Corporation
10 March 2009 (Class 704/228); filed in Japan 17 September 2003

A speech signal is transmitted with a first coded information coded in a normal state, multiplexed with second coded information used to improve quality of decoded speech when a frame loss occurs. First and second error signals are calculated from the errors between first and second target signals and synthesized signals generated by the fixed codebook. An error ratio is calculated from the ratio of the first error signal to the second error signal. A speech frame is classified according to this error ratio, and a decision is made whether or not to multiplex the second coded information for error concealment on lost packets.—DAP

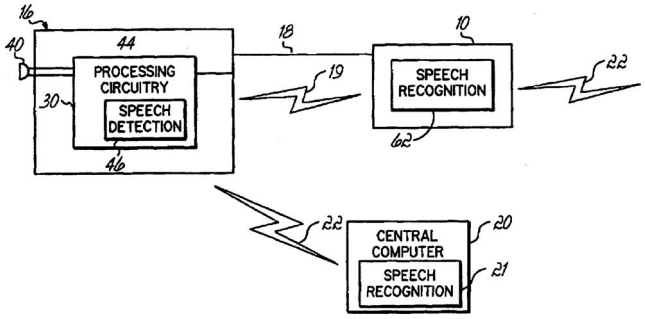


7,496,387

43.72.Ne WIRELESS HEADSET FOR USE IN SPEECH RECOGNITION ENVIRONMENT

Roger Graham Byford *et al.*, assignors to Vocollect, Incorporated
24 February 2009 (Class 455/575.2); filed 25 September 2003

In today's interconnected world a variety of tasks are performed using speech recognition systems. In many cases it would be convenient to include a wireless headset in the communication hookup. This patent points out,



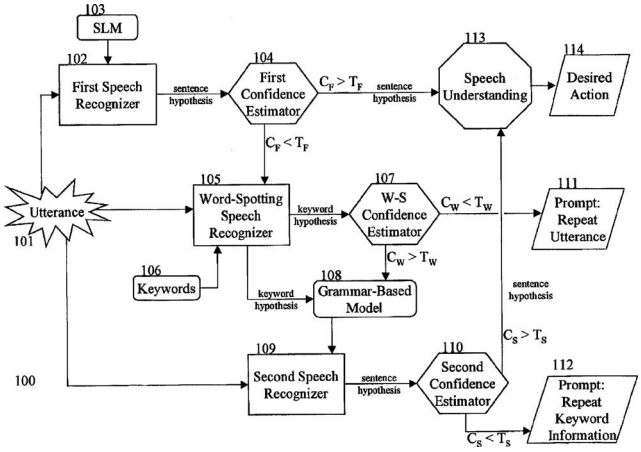
however, that a number of drawbacks are encountered with such a combination. An improved design is described which includes speech detection circuitry to avoid the need for continuous audio transmission and which transmits sampled, processed signals rather than pure audio.—GLA

7,502,737

43.72.Ne MULTI-PASS RECOGNITION OF SPOKEN DIALOGUE

Sangita R. Sharma, assignor to Intel Corporation
10 March 2009 (Class 704/251); filed 24 June 2002

First and second speech recognizers implement statistical language and grammar-based models, respectively. Confidence estimators are used on keywords for each speech recognizer to pick out utterances needing repeating if their confidence scores are below preset thresholds.—DAP



7,502,484

43.72.Ne EAR SENSOR ASSEMBLY FOR SPEECH PROCESSING

Lester S. H. Ngia *et al.*, assignors to Think-A-Move, Limited
10 March 2009 (Class 381/320); filed 14 June 2007

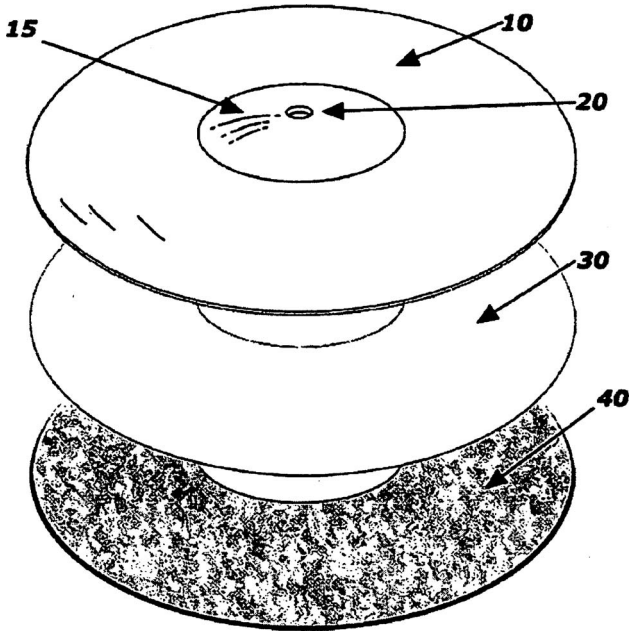
A microphone located inside the ear canal will pick up the wearer's speech but sound quality is muffled and unnatural. Nonetheless, according to this patent, the signal can be processed by speech recognition circuitry with surprising accuracy largely because of the high signal-to-noise ratio. Accuracy can be further improved by equalizing the signal to more closely match normal airborn speech.—GLA

7,408,103

43.75.Kk CYMBAL MUTING SYSTEM

Martin Richard Wachter, Baltimore, Maryland
5 August 2008 (Class 84/411 M); filed 16 May 2005

Simply put, a polyvinyl chloride annulus is glued (via adhesive disk 30) to a cymbal 20. The idea is to dampen vibrations in the cymbal by stiffening the metal plate. This is, no doubt, cheaper than using thicker metal.—MK

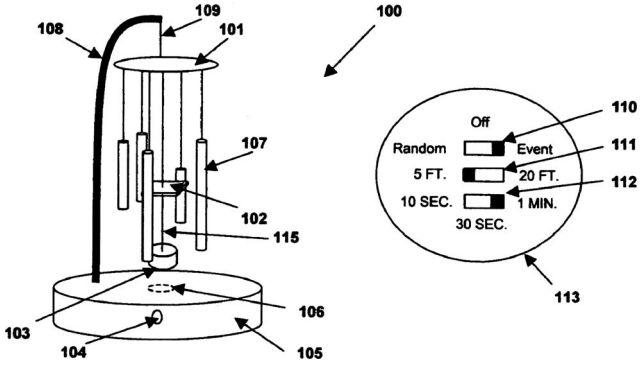


7,405,351

43.75.Tv EVENT ACTIVATED WIND CHIME SYSTEM AND METHOD OF USE

Joon Maeng, assignor to Strauss Acquisitions, L.L.C.
29 July 2008 (Class 84/404); filed 19 April 2006

Take a set of suspended tubular chimes and add either electromechanical or wind activation (possibly via a computer network) and you have a modern version of wind chimes.—MK

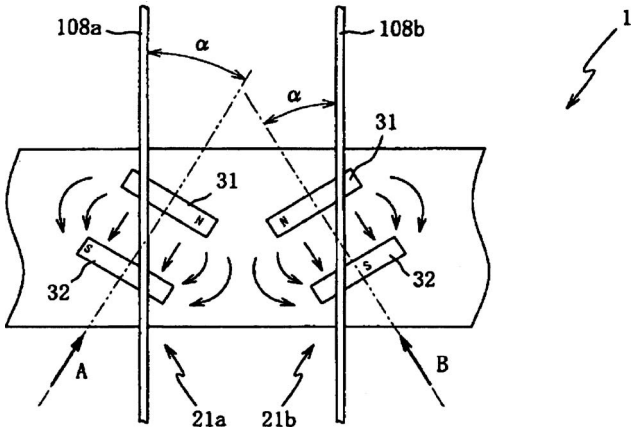


7,427,710

43.75.Tv PICKUP APPARATUS

Koji Hara, assignor to Roland Corporation
23 September 2008 (Class 84/723); filed in Japan 7 January 2005

In electric stringed instruments, the humbucker pickups exhibit two problems: (1) the adjacent string can be picked up by the transducer and (2) the pitch and amplitude can be incorrect due to the mixing of two strings. The inventor proposed placing the axes of the magnets offset by an angle. This will mitigate the crosstalk but not eliminate it, and the writing does not analyze the effect of separation and angle.—MK

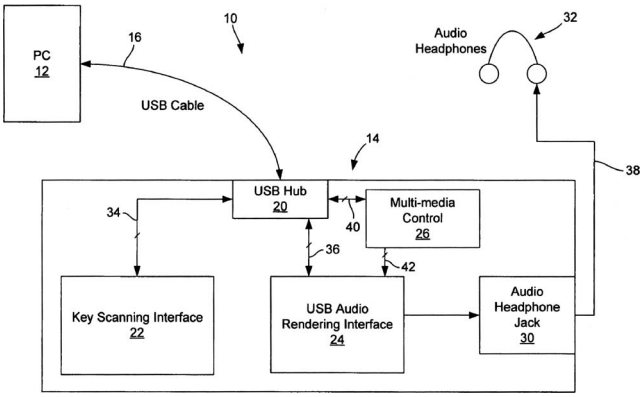


7,408,107

43.75.Wx KEYBOARD WITH AUDIO OUTPUT

Thomas H. Szolyga, assignor to Hewlett-Packard Development Company, L.P.
5 August 2008 (Class 84/615); filed 7 December 2004

To those "skilled in the art," the idea of adding an audio output to a keyboard all connected via universal serial bus is obvious. Clearly, the Patent Office thinks otherwise.—MK



7,408,109

43.75.Wx CAPACITIVE ELECTRIC MUSICAL INSTRUMENT VIBRATION TRANSDUCER

Paul John Freitas, Santa Cruz, California
5 August 2008 (Class 84/723); filed 15 February 2006

This disclosure illustrates the many and varied uses of capacitive sensors in electronic musical instruments: a drum, guitar, and cymbal are all described. Additionally, the schematics are provided for the capacitance to voltage conversion. The writing is both clear and succinct.—MK

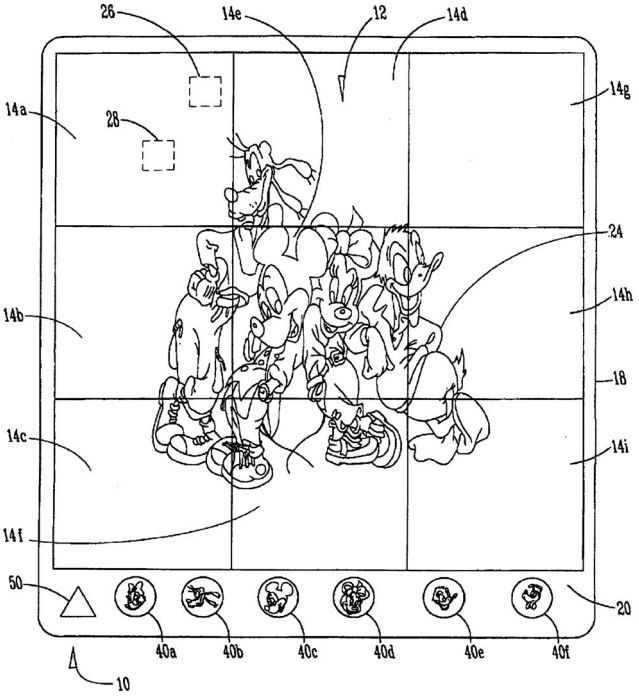
7,427,066

43.75.Wx PUZZLE APPARATUS WITH AUDIBLE SOUNDS

Richard P. Goodwin, assignor to Fenwick Enterprises LLC
23 September 2008 (Class 273/153 R); filed 19 October 2005

Imagine a set of blocks with cartoons or some other visual image. Magnets on the side of each block can activate sounds that match the char-

acter on the block. How many batteries do you need?—MK



7,494,469

43.80.Vj CONNECTION APPARATUS AND METHOD FOR CONTROLLING AN ULTRASOUND PROBE

Reinhold Bruestle, assignor to GE Medical Systems Global Technology Company, LLC
24 February 2009 (Class 600/459); filed 29 January 2004

Two chambers and a seal between them are included in an ultrasound probe. An electrical connection is made between the two chambers. The connection member in the second chamber has a flexible part and a rigid part. The rigid part forms part of the seal.—RCW

7,496,398

43.80.Vj SPATIALLY CORRELATED X-RAY AND ULTRASOUND MAMMOGRAPHIC IMAGING SYSTEMS AND METHOD

Morgan W. Nields *et al.*, assignors to Hologic Incorporated
24 February 2009 (Class 600/427); filed 27 September 2002

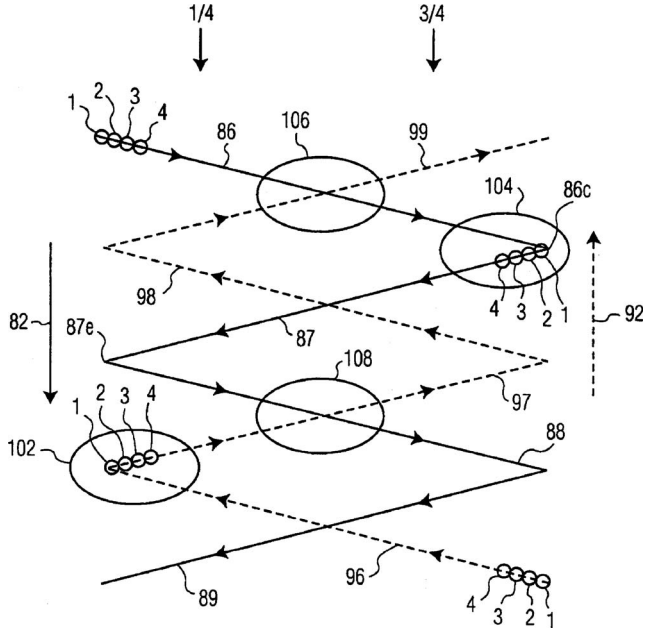
Spatially correlated x-ray and ultrasound images of the breast are obtained. The images are combined to provide three-dimensional information that locates a region of interest for analysis or biopsy. A user interface facilitates procedures such as an ultrasound-guided needle biopsy.—RCW

7,497,830

43.80.Vj THREE DIMENSIONAL ULTRASONIC IMAGING USING MECHANICAL PROBES WITH BEAM SCANNING REVERSAL

Xiang-Ning Li, assignor to Koninklijke Philips Electronics N.V.
3 March 2009 (Class 600/459); filed 2 November 2004

A one-dimensional transducer array is mechanically swept back and forth to image a volume. As the array is swept in one direction, planes are scanned in a direction that is opposite to the scan direction when the array is swept in the other direction. The result is a zigzag scanning pattern through the volume.—RCW



7,500,952

43.80.Vj PORTABLE ULTRASOUND IMAGING SYSTEM

Alice M. Chiang and Steven R. Broadstone, assignors to Teratech Corporation
10 March 2009 (Class 600/446); filed 19 July 2000

A portable battery-powered data processor and display unit is coupled by a cable to an ultrasound scan head. The scan head includes an array of transducers and electronic circuitry. The circuitry is used for transmission of ultrasound pulses and for beamformation.—RCW

7,500,954

43.80.Vj EXPANDABLE ULTRASOUND TRANSDUCER ARRAY

Walter T. Wilser and John Paul Mohr, assignors to Siemens Medical Solutions USA, Incorporated
10 March 2009 (Class 600/467); filed 22 September 2005

This transducer array folds to a smaller size for insertion and withdrawal from a patient and unfolds to provide a larger radiating surface for use in imaging a region of interest.—RCW