

Stimulus-frequency otoacoustic emission: Measurements in humans and simulations with an active cochlear model

Yong-Sun Choi and Soo-Young Lee

Brain Science Research Center and Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Republic of Korea

Kourosh Parham

Division of Otolaryngology, Department of Surgery, University of Connecticut Health Center, Farmington, Connecticut 06030, USA

Stephen T. Neely

Boys Town National Research Hospital, Omaha, Nebraska 68131, USA

Duck O. Kim^{a)}

Department of Neuroscience, University of Connecticut Health Center, Farmington, Connecticut 06030, USA

(Received 17 September 2007; revised 3 March 2008; accepted 4 March 2008)

An efficient method for measuring stimulus-frequency otoacoustic emissions (SFOAEs) was developed incorporating (1) stimulus with swept frequency or level and (2) the digital heterodyne analysis. SFOAEs were measured for 550–1450 Hz and stimulus levels of 32–62 dB sound pressure level in eight normal human adults. The mean level, number of peaks, frequency spacing between peaks, phase change, and energy-weighted group delays of SFOAEs were determined. Salient features of the human SFOAEs were simulated with an active cochlear model containing spatially low-pass filtered irregularity in the impedance. An objective fitting procedure yielded an optimal set of model parameters where, with decreasing stimulus level, the amount of cochlear amplification and the base amplitude of the irregularity increased while the spatial low-pass cutoff and the slope of the spatial low-pass filter decreased. The characteristics of the human cochlea were inferred with the model. In the model, an SFOAE consisted of a long-delay component originating from irregularity in a traveling-wave peak region and a short-delay component originating from irregularity in regions remote from the peak. The results of this study should be useful both for understanding cochlear function and for developing a clinical method of assessing cochlear status. © 2008 Acoustical Society of America. [DOI: 10.1121/1.2902184]

PACS number(s): 43.64.Jb, 43.64.Bt, 43.64.Kc [BLM]

Pages: 2651–2669

I. INTRODUCTION

Otoacoustic emissions (OAEs) are sounds generated by the inner ear (Kemp, 1978). As OAEs are related to movements of the cochlear partition, major goals of investigations of OAEs include: (1) to infer noninvasively basic characteristics of cochlear biomechanics of humans and animals in normal and pathological conditions (e.g., Kemp, 1978; Kim, 1980; Kim *et al.*, 1980; Zurek *et al.*, 1982; Kemp, 1986; Matthews and Molnar, 1986; Zwicker, 1986; Shera *et al.*, 2002; Goodman *et al.*, 2003; Siegel *et al.*, 2005) and (2) to assess the functional status of the cochlea noninvasively (e.g., Kemp *et al.*, 1990; Lonsbury-Martin and Martin, 1990; Probst *et al.*, 1991; Shera, 2004).

The present study is intended to contribute to these goals by: (1) introducing a new efficient method of measuring OAEs and providing human OAEs obtained with the method, and (2) interpreting experimentally observed OAEs in terms of an active cochlear model. Our study investigated

stimulus-frequency OAEs (SFOAEs). A presumed advantage of SFOAE over distortion product OAEs (DPOAEs) is that the information conveyed by SFOAEs may be simpler than that by DPOAEs because SFOAEs involve only one frequency whereas DPOAEs involve multiple frequencies. Most likely, though, further studies of all types of OAEs will continue to be helpful in advancing our knowledge of cochlear biomechanics and in developments of methods for assessing the cochlear functional status.

In conventional methods of SFOAE measurement (e.g., Brass and Kemp, 1993; Shera and Guinan, 1999, 2003), the SFOAE is measured for one frequency at a time. It is thus necessary to make separate measurements at a large number of frequencies if one wishes to determine high-resolution fine structures of SFOAE level and phase versus frequency.

We have developed a new efficient method for measurement of SFOAEs whereby high-resolution fine structures of SFOAE versus frequency can be obtained more rapidly. The new method employs a stimulus where frequency is continuously swept and the response waveform is analyzed using the digital heterodyne analysis. Analogously, the present method can also obtain input-output functions of SFOAE where the

^{a)}Author to whom correspondence should be addressed. Electronic mail: kim@neuron.uchc.edu.

stimulus level is continuously swept. The digital heterodyne analysis method was previously used for measurement of a time-dependent change (i.e., adaptation) of a DPOAE (Kim *et al.*, 2001) or an SFOAE (Guinan *et al.*, 2003) of a particular frequency. The latter studies involved using constant stimulus frequencies as opposed to time-varying stimulus frequencies used in the present study.

An important characteristic of SFOAEs is their fine structures in the frequency domain. Level of SFOAEs versus frequency exhibit a complex quasi-periodic pattern of peaks and troughs, while the phase of SFOAEs versus frequency also exhibits a complex pattern of many cycles of phase change indicating a frequency-dependent group delay (Kemp and Chum, 1980; Zweig and Shera, 1995; Shera and Guinan, 1999, 2003). The fine structures of SFOAEs have been interpreted to be a result of reflections of cochlear traveling waves arising from spatially distributed irregularities (or roughness) in cochlear-partition impedance (Shera and Zweig, 1991; Zweig and Shera, 1995; Talmadge *et al.*, 1998, 2000).

As an *in vivo* measurement of cochlear-partition motion is not feasible in humans, OAEs provide a useful basis for making indirect inferences for human cochlear mechanical responses. Along this line, one of the goals of the present study was to use features of human SFOAEs as a basis for specifying parameters of an active cochlear model and to make inferences about the human cochlear biomechanics. A cochlear model developed in the present study was able to reproduce salient features of the human SFOAE fine structures. Model parameters important for SFOAE simulation were the amount of cochlear amplification and the spatial-frequency contents of the partition-impedance irregularity. The correspondence between the human SFOAEs and the model results allowed us to infer characteristics of human cochlear-partition motion regarding the amount of amplification, sharpness of tuning and phase properties at various stimulus levels.

II. METHODS FOR MEASUREMENTS OF HUMAN SFOAES

The present SFOAE measurements were conducted at the University of Connecticut Health Center while Y.-S. Choi was visiting there.

A. Subjects

SFOAEs were obtained from eight ears of eight human subjects, five men and three women, aged 24–40 years. In a ninth (male, 32 years old) subject, SFOAE data were collected by the conventional method in addition to the current novel method for comparison. All of the subjects had normal hearing as documented by their audiometric thresholds being 20 dB hearing level (HL *re* American National Standards Institute S3.21, 1978) or better at octave frequencies in the range from 0.25 to 8 kHz. They also had normal patterns of tympanogram. For each subject, there were one or more recording sessions with each session typically taking 30–45 min. One session consisted of about 20 subsessions with each taking approximately 1.5 min. After groups of a

few subsessions, the subjects had short breaks (~1 min each) to allow body movements for a comfortable posture, to relieve discomfort of the probe and ear mold (see below) being in the ear canal, and to remove any potential long-term adaptation.

During the recording sessions, the subjects remained awake sitting quietly on a chair in a double-walled acoustic chamber (Industrial Acoustics Co.). The present study was approved by the Institutional Review Board for Human Subjects of University of Connecticut Health Center.

B. Equipment

Stimuli were synthesized, stimulus presentation was controlled, and acoustic waveforms were recorded using a Windows-PC which contained a sound card (Card Deluxe, Digital Audio Labs) with 24 bit analog-to-digital and digital-to-analog converters operating at a sampling rate of 22 kHz. We used a program called SYSRES (Neely and Stevenson, 2002) to present the custom-synthesized stimulus and collect the acoustic waveforms.

The acoustic system included one microphone and two earphones (ER-10C, Etymotic Research) designed for ear-canal insertion. Individualized ear molds were made for the human subjects with silicon-based impression material. During the recording period, the ear mold was inserted into a subject's ear canal, and the tube of the ER-10C acoustic probe was inserted into the hole running through the ear mold. This arrangement, having a tight coupling between the acoustic probe and the ear canal, was intended to provide a higher signal-to-noise ratio and a more controlled positioning of the probe than a conventional arrangement where the probe was inserted in the ear canal with a foam tip.

C. Suppression method to extract SFOAEs

The measurement of SFOAEs requires separation of the recorded acoustic signal into the “stimulus” and emission components. The stimulus component represents both a signal generated by an earphone and passive acoustic properties of the earphone coupler and the passive acoustic input impedance looking into the ear canal. The emission component represents a signal generated by the ear. One method of such separation is based on the fact that an SFOAE is suppressed when a suppressor tone is added at a frequency near the probe tone (e.g., Kemp and Souter, 1988; Brass and Kemp, 1993; Shera and Guinan, 1999). In this method, the stimulus component is obtained from a condition where both a probe tone and a suppressor tone are applied as the stimulus. An SFOAE is then obtained by vector subtraction of the stimulus component from the signal containing both the stimulus and SFOAE. The present study used this method.

D. Frequency-sweeping method: Stimulus generation

The stimulus frequency was continuously swept from 500 to 1500 Hz in order to obtain high-resolution fine structures of SFOAEs. We applied a stimulus signal and recorded pressure response of the ear canal. At each measurement, the acoustic stimulus was repeated 32 times. As the duration of the stimulus was approximately 3 s, it took about 1.5 min for

TABLE I. Midway stimulus level in dB SPL and the range of levels obtained at the input attenuation level of -50 dB *re* maximum voltage into the earphone for each subject. The last column shows the corresponding information pooled for all subjects.

| Subject ID No. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | all |
|--------------------|----|----|----|----|----|----|----|----|-----|
| Mid-level (dB SPL) | 49 | 39 | 39 | 37 | 42 | 40 | 37 | 39 | 42 |
| Range (\pm dB) | 2 | 3 | 7 | 4 | 3 | 8 | 4 | 3 | 10 |

a measurement covering 500–1500 Hz at one stimulus level. The stimulus had two channels delivered through two channels of the soundcard which were connected to two earphones of the ER-10C probe. One channel of the stimulus produced a probe tone whereas the other channel produced a suppressor tone. Each channel had two equal-length (32768 samples/(22000 samples/s=1.489 s) intervals. The two intervals of the probe channel were identical whereas the first interval of the suppressor channel was silence.

The probe and suppressor signals, $p(t)$ and $s(t)$, were:

$$p(t) = A_p \cdot \sin \left[2\pi \cdot \left\{ 500 \times \frac{f_{\text{sam}}}{31448} \times (t - 0.03)^2 + 500 \times (t - 0.03) \right\} \right], \quad (1)$$

$$s(t) = A_s \cdot \sin \left[2\pi \cdot \left\{ 500 \times \frac{f_{\text{sam}}}{31448} \times (t - 0.03)^2 + 650 \times (t - 0.03) \right\} \right]. \quad (2)$$

A_p and A_s represent amplitudes of the probe and suppressor tones, respectively, f_{sam} the sample rate in Hz. The unit of time, t , was seconds (s). The above equations included 0.03 s silent portions that were intended to avoid transients. The probe frequency was swept linearly from 500 to 1500 Hz. The suppressor frequency was also swept linearly such that it was 150 Hz above the probe frequency. For both $p(t)$ and $s(t)$, we used 30 ms rise/fall times with half Blackman windows. After subtracting the silent and rise/fall portions, the above signal waveform was stable for 1.37 s.

We converted the voltage amplitude of a stimulus tone to the earphone, A_p or A_s , relative to the maximum voltage at attenuation of the tone in dB. A constant attenuation corresponded to a variable sound pressure level (SPL) as a result of a frequency dependence of the earphone and the acoustic load impedance of the ear canal. In the present study, -50 dB attenuation, for example, corresponded to 37 ± 4 dB SPL (*re* 20 μ Pa) for 500–1500 Hz in one subject (No. 7). Further details on this topic are described in Sec. III A, Table I.

For each measurement, the probe stimulus was applied at a constant attenuation with the probe frequency swept from 500 to 1500 Hz. The attenuation level of the probe tone varied from -70 dB to -25 dB in 5 dB steps. In the present report, only the results with attenuation of -60 dB to -30 dB in 10 dB steps were systematically analyzed across all subjects. The suppressor level was 20–40 dB above that of the probe level; i.e., the suppressor level was -20 , -15 , -10 , and -5 dB *re* max when the probe level was -60 , -50 , -40 , and -30 dB *re* max, respectively. When the probe level was

higher, we had to use smaller difference between the suppressor level and the probe level in order to keep the absolute level of the suppressor.

E. Digital heterodyne analysis of SFOAEs

We used the digital heterodyne method to determine SFOAE level and phase as a function of frequency which, in turn, are functions of time in the present study. The digital heterodyne method, a Fourier-transform application, was developed by Stephen T. Neely (Boys Town National Research Hospital, Omaha, NE, personal communication), and previously used to measure an OAE at a particular frequency as a function of time for a fixed stimulus (Kim *et al.*, 2001; Guinan *et al.*, 2003).

In the context of the present study, we took a discrete Fourier transform of the 1.43-s-long original response waveform sampled at 22 kHz. A band of positive frequencies (1000 spectral components) around an SFOAE analysis frequency was then selected in the frequency domain and shifted down (“heterodyned”) such that the SFOAE frequency was set to be the zero frequency. Next, a Blackman low-pass filter was applied to the shifted spectrum.

The optimal bandwidth (“BW”) of the heterodyne low-pass filter is proportional to square root of the frequency sweep rate. This is because opposing constraints on the BW are imposed in the frequency- and time-domain parts of the heterodyne analysis. An unduly wide BW of the low-pass filter will allow energy at a frequency remote from an analysis frequency to be included in the frequency-domain part of the heterodyne analysis whereas an unduly narrow BW of the filter (which has a long integration time), together with a frequency-swept stimulus, will allow energy at a frequency remote from the analysis frequency to be included in the time-domain part of the heterodyne analysis. We empirically found, for a Blackman low-pass filter used in this study, the optimal value of BW (defined as the bandwidth at a relative amplitude of 0.7):

$$\text{BW} = 0.345 * (\text{sr})^{0.5}, \quad (3)$$

where sr corresponded to sweep rate in Hz/s. [Note: the unit of $(\text{Hz}/\text{s})^{0.5}$ is Hz.] By substituting $\text{sr}=699.3$ Hz/s (1000 Hz/1.43 s) into Eq. (3), we obtained $\text{BW}=9.1$ Hz.

We then took an inverse discrete Fourier transform of the shifted and low-passed, yielding a complex-valued time-domain signal. The latter consisted of 1000 time points representing the magnitude and phase of the SFOAE of a particular analysis frequency as a function of time over 1.43 s with each time point representing 1.43 ms.

Originally, the heterodyne method was used to extract one specific frequency component as a function of time for a

fixed stimulus. A novel aspect of the present method is that a time dependence of the heterodyne analysis output is made to be equivalent to a frequency dependence by combining a frequency-swept stimulus with the heterodyne analysis.

To extract an SFOAE at a particular analysis frequency, the total 1.43 s recorded signal was first multiplied with a 227 ms Blackman window (spanning a frequency range of ± 79 Hz) centered at the point when the probe-tone frequency was equal to the analysis frequency to suppress the suppressor-tone signal; the suppressor frequency was 150 Hz higher than the probe frequency. The magnitude of the heterodyne-analysis output for a particular analysis frequency showed a peak near a time when the stimulus frequency was equal to the analysis frequency. A peak occurred slightly later (at a time to be called “ t_{peak} ”) than the time when the stimulus frequency was equal to the analysis frequency (“ t_{stimulus} ”). The average delay between t_{peak} and t_{stimulus} among all of the present data was 3.1 ms. This delay represents a combination of a round-trip delay of the SFOAE signal and a delay of the measurement system.

Each original acoustic waveform was shifted by 3.1 ms to compensate for the delay described above. Then the heterodyne analysis was applied to the shifted waveform, and the magnitude, $m(f)$, and phase, $\theta(f)$, of an SFOAE for a particular analysis frequency, f , was determined by taking the magnitude and phase of the heterodyne-analysis output at the time when the stimulus frequency was equal to the analysis frequency. We converted the SFOAE magnitude, m in μPa , into an SFOAE level in dB SPL *re* 20 μPa by using

$$\text{SFOAE level} = 20 \cdot \log_{10}(m/20 \mu\text{Pa}). \quad (4)$$

We repeated the above heterodyne analysis for a particular response waveform with a large number of analysis frequencies ranging from 550 to 1450 Hz in small steps. Because the analysis could be performed after the data were collected, the analysis time did not increase the recording time. The frequencies of 500–550 and 1450–1500 Hz were excluded to avoid transients. The minimum frequency separation where the contribution to an analysis frequency was less than 0.7 relative to the contribution by the analysis frequency itself was given by the bandwidth of the low-pass filter, ± 9.1 Hz. Thus, for the present SFOAE measure at each analysis frequency, the signal component at a frequency 9.1 Hz away from the analysis frequency had a relative contribution of 0.7. In describing our results, we chose to display the SFOAE measures versus frequency with a 5 Hz spacing rather than a 9.1 Hz spacing to improve visualization.

F. Noise floor computation

Noise floor was computed in separate steps from the normal SFOAE analysis. As stated above, the heterodyne output corresponds to the level and phase of a particular analysis frequency versus time; the full time range of measured waveform (0–1.43 s) in the present study was mapped to a probe-stimulus-frequency range of 500–1500 Hz. Such a heterodyne output typically exhibited high levels near the points where the probe frequency or suppressor frequency (150 Hz above the probe frequency) was equal to the analy-

sis frequency. The level near the point of suppressor frequency was attenuated by applying a 227 ms Blackman window (spanning a frequency range of ± 79 Hz) centered at the probe frequency, as stated above. In a wide region of the full range of time/frequency in the heterodyne analysis output, neither the probe frequency nor the suppressor frequency was near the analysis frequency. In such a “remote region,” consequently, the level of the heterodyne analysis output was quite low.

We defined a remote region to be a region spanning probe frequencies over a 400 Hz range. For this purpose, we treated the full probe frequency range of 550–1450 Hz as a circular range. For a particular analysis frequency, a probe frequency in the remote region was at least 150 Hz higher than, and at least 350 Hz lower than, the analysis frequency. For example, if the analysis frequency was 1000 Hz, the remote region consisted of two subregions of probe frequencies: (1) 550–650 Hz, and (2) 1150–1450 Hz.

For a particular analysis frequency, the 400-Hz-wide remote probe-frequency region consisted of 80 sample points separated by 5 Hz steps. For each of the 80 remote probe-frequency points, we multiplied the original response waveform with a 227 ms Blackman window centered at the “remote point” under consideration and then performed the heterodyne analysis yielding one remote level. We repeated the process 80 times for the 80 remote points and defined the average of the 80 remote levels to be the noise floor level for one analysis frequency under consideration. This process was repeated for each of all analysis frequencies.

G. Quantification of response features of SFOAE fine structures

SFOAEs exhibited fine structures, i.e., multiple peaks and troughs of level and many cycles of phase change over a frequency range of 550–1450 Hz. Based on such data in each subject, we quantified the following response features: (1) mean level (ML) of SFOAE in dB SPL; (2) number of peaks (NP); (3) frequency spacing (Δf) between SFOAE peaks in Hz; (4) number of cycles (NC) of SFOAE phase change; (5) energy-weighted group delay (EGD) in ms; and (6) standard deviation of EGD (σ_{EGD}) in ms.

Regarding NP, we required that an acceptable peak should be surrounded troughs each with a minimum 2 dB depth; the latter was measured between a relevant local minimum and the straight line connecting two neighboring peak candidates (local maxima). To determine energy-weighted group delays (EGDs), we first unwrapped SFOAE phase versus frequency where the neighboring phase values were constrained to be within one half cycle. The EGD was obtained by differentiating the unwrapped SFOAE phase, θ , with respect to frequency, f , with an energy weighting as follows:

$$\text{EGD}(f_c) = \frac{-\sum_f A^2(f) \cdot \frac{d\theta}{df}}{\sum_f A^2(f)}; \quad (5)$$

where, $A(f)$ and f_c represent the SFOAE amplitude at frequency f and the center of the processed frequency range,

respectively. The value of $d\theta/(df)$ was obtained by dividing the difference between the two neighboring phase values by the 5 Hz frequency difference. We used 25-Hz-wide ranges centered at various frequencies for information about how a narrowband EGD changes as a function of the center frequency, and a 900-Hz-wide range for information about the overall EGD processed for the whole 550–1450 Hz range. This definition of EGD was adapted from Goldstein *et al.* (1971) who indicated that EGD of a band-limited filter corresponds to the center of gravity of the impulse response of the filter. The advantage of EGD over the un-weighted group delay (GD) associated with a large amplitude is more meaningful than one associated with a small amplitude. We also determined normalized energy-weighted group delays (NEGDs) in units of cycles as follows:

$$\text{NEGD}(f_c) = \text{EGD} \times f_c. \quad (6)$$

The measure σ_{EGD} describes the variability of 25-Hz-wide EGDs over the 550–1450 Hz range in each subject.

H. Level-sweeping method

The level-sweeping method allowed us to measure the SFOAE input-output functions at specific frequencies. The frequency-swept data obtained at -40 dB *re* maximum were used to choose several peak and trough frequencies. Measurements were made at selected probe frequencies with swept stimulus levels. After the measurement, the probe (f_1) component was obtained using the heterodyne analysis method as above. The input-output functions of the SFOAE level and phase at a specific probe frequency were constructed by pairing the output level and phase with the corresponding input level.

At each measurement, the acoustic stimulus was repeated eight times. The duration of the stimulus was approximately 12 s hence each measurement required about 1.5 min. The stimulus had two channels with each having two intervals (as above). The probe frequency was fixed but level was varied, and the suppressor frequency was fixed at 150 Hz above the probe frequency and the level was varied. During the level sweep, the amplitudes of the waveforms were exponentially increased or decreased. These exponential changes of amplitudes over time correspond to linear sweeps of the stimulus levels in dB SPL versus time.

To construct the SFOAE input-output function, the heterodyne analysis was applied to extract the probe frequency component. Finally, the SFOAEs were extracted by taking the vector difference between the two intervals (as above). Generally, the probe level ranged from 10 to 70 dB SPL with the suppressor level ranging from 50 to 90 dB SPL; the suppressor level was 20–40 dB above the probe level. In cases where the maximum ear canal pressure of the suppressor channel was less than 90 dB SPL, the ranges of the probe and suppressor levels were adjusted. The present method of obtaining input-output functions with a continuously swept stimulus level is similar to that described by Neely *et al.* (2003).

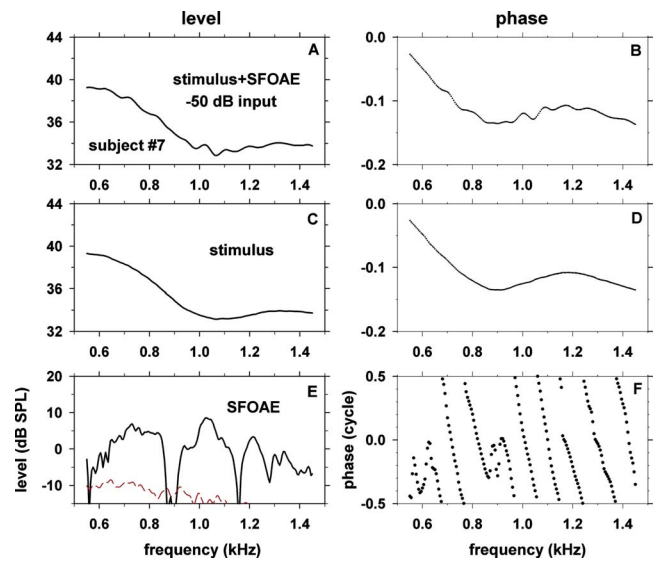


FIG. 1. (Color online) Procedure for obtaining SFOAE level and phase using frequency-swept stimuli illustrated with results obtained with a stimulus level of -50 dB *re* maximum voltage in one subject, No. 7. The left and right columns show signal level (in dB SPL) and phase (in cycles), respectively. In the top row (panels A and B), the signal (in the first interval) was a mixture of stimulus and SFOAE. In the second row (panels C and D), the signal (in the second interval) was stimulus with little SFOAE present as SFOAE was suppressed by the suppressor tone. The vector difference between the signals in the top two rows, shown in the third row (panels E and F), corresponds to SFOAE. Negative phase values were defined to be a phase lag. In panel E, dashed line represents noise floor.

III. RESULTS OF HUMAN SFOAES

A. Frequency-sweeping experiment

The procedure for obtaining SFOAE level and phase versus frequency is illustrated with an example obtained in one subject (No. 7) with a probe stimulus level of -50 dB *re* maximum voltage into the earphone (Fig. 1). As described in Methods, the stimulus consisted of two intervals where the first interval contained only the probe frequency component and the second interval contained both the probe and the suppressor frequency components. We obtained SFOAE level and phase versus frequency (Fig. 1, bottom row) by taking vector differences between the heterodyne outputs derived from the first interval (corresponding to SFOAEs plus probe stimuli, Fig. 1, top row) and those derived from the second interval (corresponding to probe stimuli, Fig. 1, middle row). The signals in the middle and bottom rows of Fig. 1 display a slowly varying pattern superimposed on a rapidly varying pattern. The slowly varying stimulus (middle row) represents the frequency-dependent properties of the stimulus-generating system and the input impedance looking into the ear. The dashed line in panel E shows the noise floor computed as described in Methods.

The data of Fig. 1 are consistent with the concept that the stimulus+SFOAE fine structure has a frequency spacing the reciprocal of which corresponds to a round-trip delay (Zweig and Shera, 1995). The frequency spacing between peaks of the stimulus+SFOAE around 1 kHz [Fig. 1(A)] is 81 Hz. The reciprocal of 81 Hz is 12.3 ms. This value is close to the group delay of SFOAEs around 1 kHz, 11.1 ms [see below, Fig. 2(I)].

TABLE II. Comparison of SFOAEs obtained with the conventional analysis method against those obtained with the heterodyne analysis method. The two types of analysis methods were applied to a common set of data records obtained with conventional stimuli where the stimulus frequency was kept constant (at one of three values) over a 1.43 s stimulus period. The stimulus levels were 57, 54, and 52 dB SPL for 800, 1000, and 1100 Hz, respectively. Each entry is level (in dB SPL) and phase (in cycle) of SFOAE for a particular frequency. Results from one ear of Subject No. 9.

| | 800 Hz | 1000 Hz | 1100 Hz |
|--------------|---------------------|---------------------|---------------------|
| Heterodyne | 24.8 dB; 0.511 cyc. | 23.6 dB; 0.480 cyc. | 22.3 dB; 0.038 cyc. |
| Conventional | 24.5 dB; 0.524 cyc. | 23.3 dB; 0.482 cyc. | 22.2 dB; 0.035 cyc. |

As seen in panel C of Fig. 1, ear-canal sound pressure level varied as a function of frequency when the stimulus level was kept at a constant attenuation of voltage into the earphone. In this subject, -50 dB *re* maximum voltage corresponded to 37 ± 4 dB SPL *re* $20 \mu\text{Pa}$. The range of dB SPL values in various subjects at -50 dB *re* maximum voltage are described in Table I.

To verify that the current method generates results consistent with those of the conventional method (using a fixed-frequency stimulus), we analyzed a common dataset obtained with a constant stimulus frequency using two analysis methods: (1) a simple conventional Fourier transform, and (2) the heterodyne analysis. The results of one subject (No. 9) for three stimulus frequencies are summarized in Table II. The differences between the results obtained by the two analysis methods were within 0.3 dB in level and 0.013 cycle in phase. These data demonstrate that the results obtained with the current heterodyne analysis method are consistent with those obtained with the conventional method.

The results obtained with stimulus level from 27 to 57 dB SPL in subject No. 7 are shown in Fig. 2. The three columns represent SFOAE level, phase and 25-Hz-wide

EGDs, respectively. The effective frequency resolution of the present data is 9.1 Hz as stated in Methods, and the data are plotted with 5 Hz steps. Prominent features of the data are: (1) SFOAE level versus frequency exhibits many quasi-periodic peaks and troughs; (2) when SFOAE phase is unwrapped, the phase versus frequency tends to be almost a straight line with a steep slope but there are conspicuous irregularities in the slope; (3) the 25-Hz-wide EGD varies irregularly as a function of frequency; and (4) all of the SFOAE fine-structure features change noticeably with stimulus level. The quasi-periodic peaks/troughs in SFOAE level versus frequency (Fig. 2, left column) should not be equated to quasi-periodic peaks/troughs seen in a mixture of SFOAE and stimulus [Fig. 1(A)]. An SFOAE itself (after removing the stimulus) still exhibits a quasi-periodicity. The mean level (ML) of SFOAE changed from -7.2 dB SPL at stimulus level of 27 dB SPL to 11.6 dB SPL at stimulus level of 57 dB SPL. The number of peaks (NP) of SFOAE in the 550–1450 Hz range decreased from 21 peaks at 27 dB SPL to six peaks at 57 dB SPL. The frequency spacing between peaks, Δf , increased from 42 to 121 Hz and the number of cycles of phase change (NC) decreased from 8.0 to 6.0 cycles over the same range of stimulus levels.

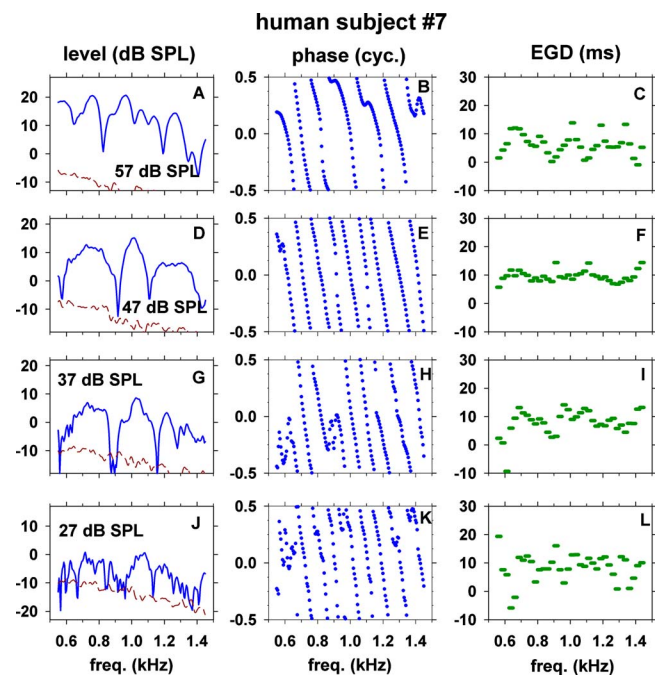


FIG. 2. (Color online) Level, phase, and 25-Hz-wide energy-weighted group delay (EGD) of SFOAE in one subject (No. 7) versus frequency at several stimulus levels (27–57 dB SPL). Dashed lines in the left panels show noise floor.

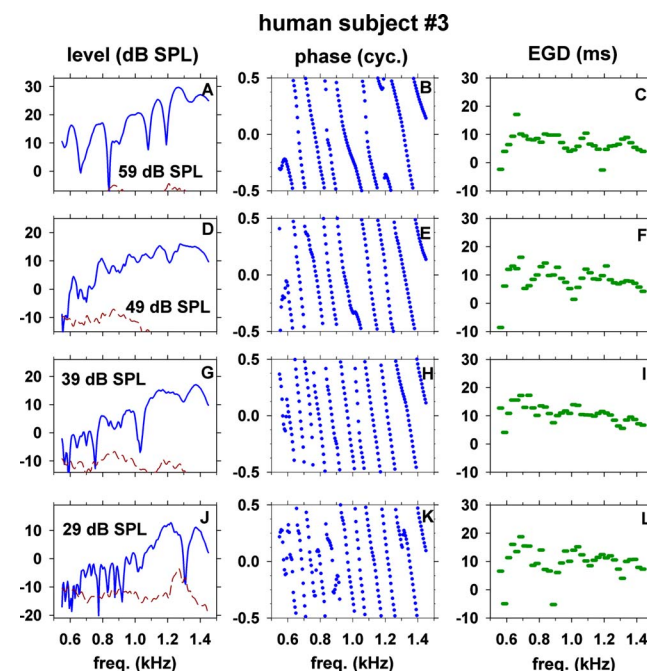


FIG. 3. (Color online) Level, phase, and 25-Hz-wide EGD of SFOAE in another subject (No. 3) versus frequency at several stimulus levels (29–59 dB SPL).

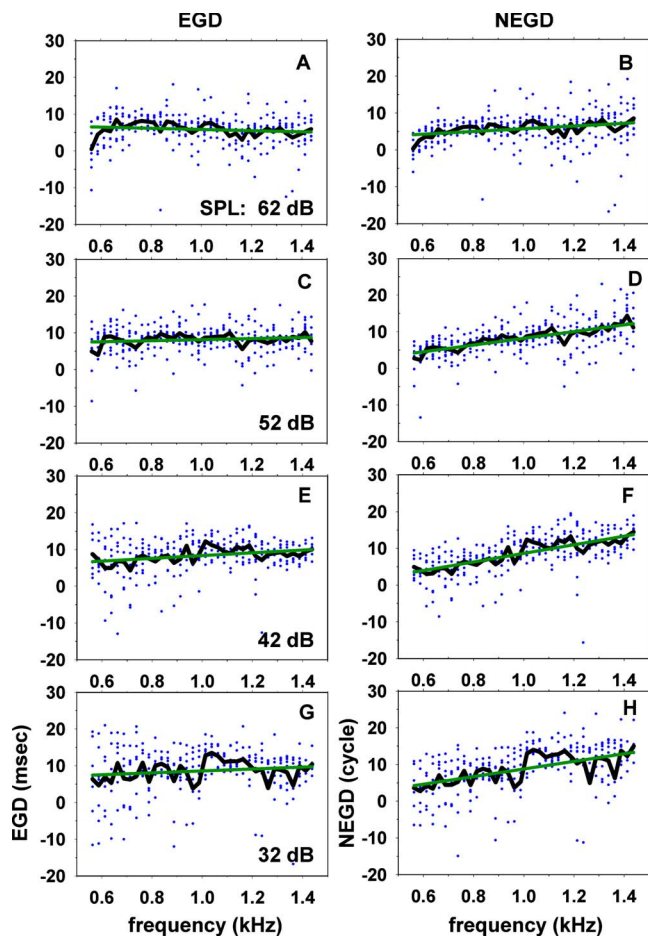


FIG. 4. (Color online) Energy-weighted group delays (EGDs) and normalized EGDs (NEGDs) of all subjects at several stimulus levels (32–62 dB SPL). Individual dots represent 25-Hz-wide EGDs or NEGDs of all subjects. At every 25 Hz band, individual EGDs or NEGDs were averaged and shown as a jagged line in each panel. A straight line is a linear regression line for the average values for various frequencies.

Results from another subject (No. 3) are shown in Fig. 3 for stimulus levels of 29–59 dB SPL to illustrate an inter-subject variability. Although there are quantitative differences between the two subjects, the results from the two subjects are qualitatively similar; e.g., (1) there are many peaks/troughs of SFOAE level, (2) there are many cycles of phase change, (3) the 25-Hz-wide EGD varies irregularly as a function of frequency; and (4) NP, Δf , and NC change with stimulus level. ML changed from -0.5 dB SPL at stimulus level of 29 dB SPL to 17.5 dB SPL at stimulus level of 59 dB SPL. NP decreased from 16 to 7 peaks, Δf increased from 54 to 133 Hz and NC decreased from 9.2 to 6.6 cycles over the range of 29–59 dB SPL stimulus level.

Major troughs of SFOAE level tended to have concomitant rapid changes in phase slope and EGD. For example, there were major troughs at ~ 820 , 1190, and 1410 Hz in Fig. 2(A) (57 dB SPL) and at ~ 910 and 1110 Hz in Fig. 2(D) (47 dB SPL). At such frequencies, the phase slopes abruptly changed [Figs. 2(B) and 2(E)] and the EGD values deviated from the neighboring values [Figs. 2(C) and 2(F)]. The phase slope became either steeper, shallower, or inverted with the EGD becoming longer, shorter, or negative. Analogous examples of concomitant microstructures of level, phase, and

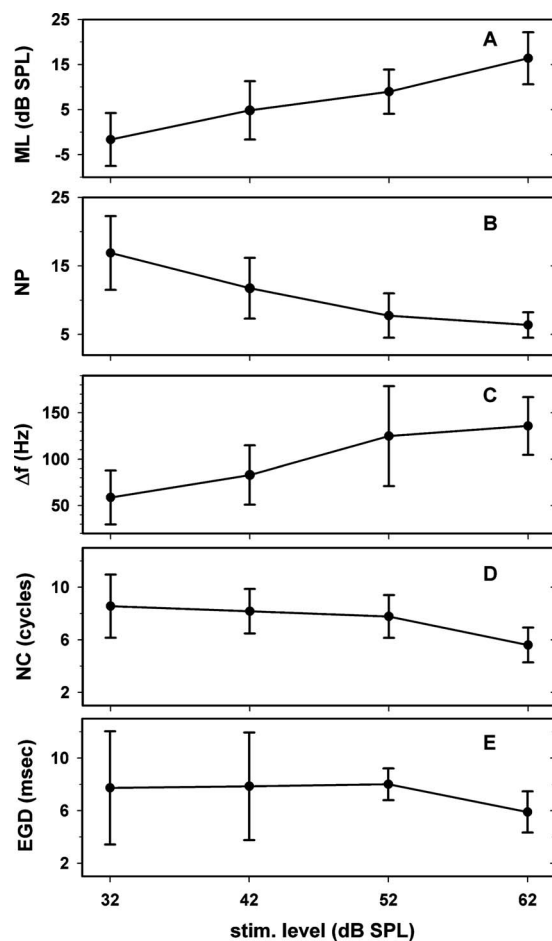


FIG. 5. Means \pm standard deviations of five response features of SFOAEs among eight human subjects versus stimulus level: mean level (ML) of SFOAE, number of peaks (NP), frequency spacing (Δf), number of cycles of phase change (NC), and energy-weighted group delays (EGDs) in ms.

EGD of SFOAEs were also observed in another subject, e.g., troughs at ~ 1080 and 1200 Hz [Fig. 3(A)] were associated with rapid or irregular changes of phase slope [Fig. 3(B)] and EGD [Fig. 3(C)].

The 25-Hz-wide EGDs and normalized EGDs (NEGDs) of all eight subjects are plotted versus frequency as scattered dots in Fig. 4. The jagged and straight lines represent the average values and linear regression lines of the average values, respectively. There was a large variability in EGD and NEGD across subjects and across frequency. However, the linear regressions showed that EGDs tended to be almost constant and NEGDs tended to increase with increasing frequency across the 550–1450 Hz range. The value of EGD tended to decrease with increasing stimulus level. For example, EGD near 1 kHz tended to be ~ 9 ms at 32 dB SPL and ~ 6 ms at 62 dB SPL; the corresponding values of NEGD were 9 at 32 dB SPL and 6 at 62 dB SPL.

We determined ML, NP, Δf , NC and EGD for all subjects. Means and standard deviations of each of these measures across the subjects are shown at four stimulus levels in Fig. 5. For this figure, EGDs were calculated over the full 900-Hz-wide range NP, NC, and EGD decreased whereas ML and Δf increased with increasing stimulus level.

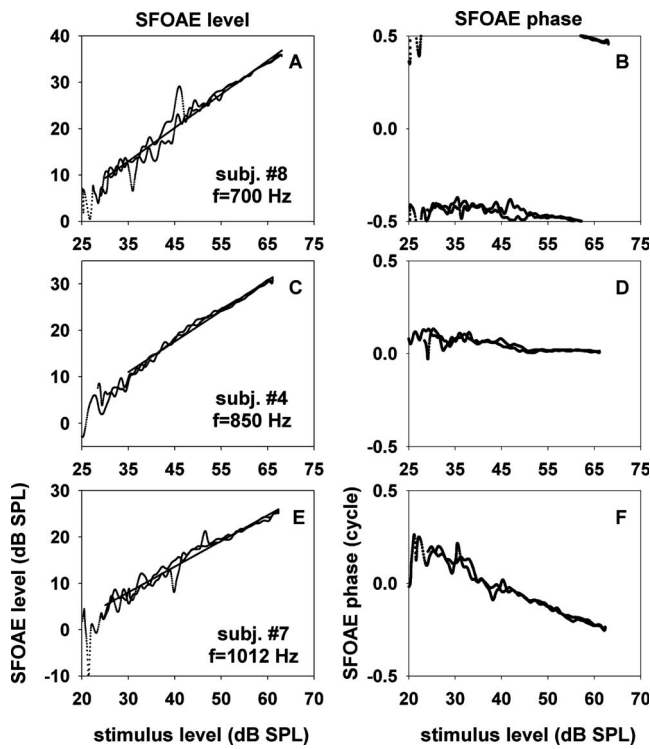


FIG. 6. SFOAE level and phase versus stimulus level for “peak stimulus frequencies.” The latter were defined to be stimulus frequencies corresponding to peaks of SFOAE level versus frequency at 52 dB SPL. The three rows correspond to data from three different subjects; the subject identification numbers and the selected stimulus frequencies are shown in the insets. The input-output functions of SFOAE level were fitted with straight lines. The slopes of the fitting lines were 0.72, 0.66, and 0.55 dB/dB for panels A, C, and E, respectively.

B. Level-sweeping experiment

We determined dependence of SFOAE level and phase upon stimulus level by using the stimulus level-sweeping method described in Sec. II H. For these measurements, we selected the probe frequency corresponding to a peak or a trough (in a graph like Figs. 2 and 3). Results obtained with a peak frequency in three subjects are shown in Fig. 6. The input-output functions in log-log plots were approximately straight lines. The data of Fig. 6 were fit with slopes ranging from 0.55 to 0.72 dB/dB. The phase of SFOAE remained nearly independent of stimulus level or became lagging as the stimulus level increased. The amount of phase change was small (<0.14 cycle) in two (panels B and D) of the three cases and larger (0.47 cycle) in the third case (panel F). Slope and phase change obtained with a peak frequency in each subject are described in Table III. The input-output slope ranged 0.52–0.72 whereas the amount of phase change ranged -0.13 to -0.47 ; a negative phase change corresponds to a phase lag.

TABLE III. Slopes of input-output functions of SFOAE level and changes of SFOAE phase with stimulus level at frequencies corresponding to SFOAE-level peaks.

| Subject ID No. | 2 | 3 | 4 | 5 | 7 | 8 |
|------------------------|-------|-------|-------|-------|-------|-------|
| Slope (dB/dB) | 0.52 | 0.66 | 0.66 | 0.55 | 0.55 | 0.72 |
| Δ phase (cycle) | -0.25 | -0.26 | -0.13 | -0.47 | -0.45 | -0.14 |
| Frequency (Hz) | 820 | 961 | 850 | 1075 | 1012 | 700 |

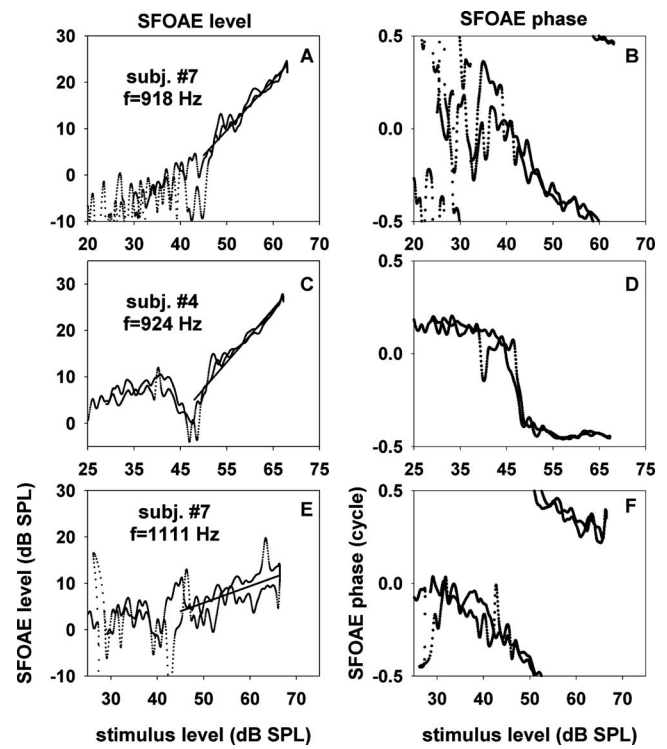


FIG. 7. SFOAE level and phase versus stimulus level for “trough stimulus frequencies.” The latter were defined to be stimulus frequencies corresponding to troughs of SFOAE level versus frequency at 52 dB SPL. The three rows correspond to data from three different cases; the subject identification numbers and the selected stimulus frequencies are shown in the insets. The portions of the input-output functions of SFOAE level, where slopes could be ascertained, were fitted with straight lines. The slopes of the fitting lines were 1.08, 1.15, and 0.36 dB/dB for panels A, C, and E, respectively.

Analogously, we determined dependence of SFOAE level and phase upon stimulus level for trough frequencies as shown in Fig. 7. These data illustrate that SFOAE versus stimulus level can be complex and variable if the probe frequency falls in a trough of SFOAE level versus frequency. The example of panel C is a clear illustration of a nonmonotonic input-output level function together with a rapid phase change of about a half cycle around the level notch. The present level-sweep method is particularly useful in characterizing such complex input-output functions because a high resolution of stimulus level is necessary to determine the nonmonotonic nature of emission level and a rapid phase change restricted to a small region of stimulus level.

In the left panels of Fig. 7, partial regions were selected where slopes could be measured. These slopes are summarized in Table IV. Two entries for a particular subject correspond to results obtained with two trough frequencies. The input-output slopes for trough frequencies ranged from 0.36 to 1.83. The range of slopes for trough frequencies was wider

TABLE IV. Slopes of input-output functions of SFOAE level at frequencies corresponding to SFOAE-level troughs in regions where slopes could be defined. Multiple entries within a subject represent multiple trough frequencies. There was no correlation between slope and frequency.

| Subject ID No. | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|----------------|------|------|------|------|------|------|------|
| Slope (dB/dB) | 0.50 | 0.37 | 0.87 | 0.61 | 0.48 | 0.36 | 0.75 |
| Frequency (Hz) | 1320 | 896 | 1121 | 1420 | 1200 | 1111 | 1450 |
| Slope (dB/dB) | NA | 1.83 | 1.53 | 0.75 | 0.75 | 1.08 | 0.78 |
| Frequency (Hz) | NA | 1014 | 1193 | 1295 | 1396 | 918 | 770 |

than that for peak frequencies. Measurements of phase changes in this case were omitted because the starting or ending points of the phases were not clear in most cases.

The facts that SFOAE level can change nonmonotonically with stimulus level and that the peak frequencies shift with stimulus level are illustrated with results from one subject in Fig. 8. It describes SFOAE level over a small range of frequencies (1000–1250 Hz) around a peak and the neighboring troughs for various stimulus levels. At 1160 Hz, SFOAE level was a nonmonotonic function of stimulus level similar to Fig. 7(C). The peak of SFOAE level was centered at 1140 Hz at stimulus level of 22 dB SPL but it gradually shifted toward lower frequencies with increasing stimulus level. The peak frequency decreased to 1065 Hz at 62 dB SPL.

IV. MODEL DEFINITION AND METHODS

A. Basic concept

The present model is intended to represent the human cochlea with a length of 35 mm. To simulate SFOAEs one needs to calculate the forward and backward waves in the cochlea by solving the wave equation. We adopted the one-dimensional transmission line model of the cochlea as described by [Zweig et al. \(1976\)](#). The nonuniform transmission line model incorporates a second-order differential equation for the wave propagation and reflection as

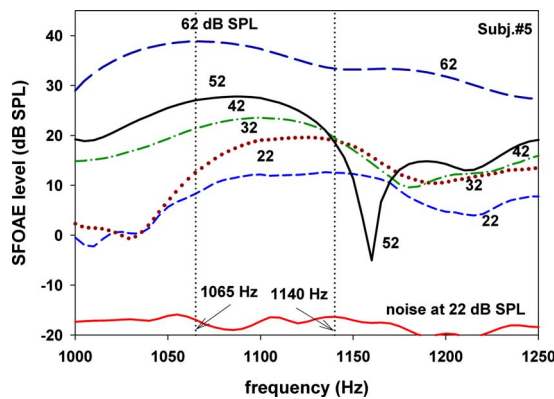


FIG. 8. (Color online) An example illustrating that a peak of SFOAE level shifted as a function of stimulus level. As the stimulus level increased, the position of the peak shifted to a lower frequency. The data also illustrate the fact that SFOAE level for a frequency near a trough (1160 Hz) can be a nonmonotonic function of stimulus level. The lowest curve shows noise floor at 22 dB SPL.

$$\frac{d^2P}{dx^2} - \frac{d \ln Z_s}{dx} \frac{dP}{dx} - Y_b Z_s P = 0, \quad (7)$$

where P is pressure difference between scalae, Z_s scala impedance per centimeter, Y_b cochlear-partition admittance per unit length, and x distance from the stapes. The cochlear-partition admittance and scala impedance were selected to maintain an approximate shift invariance. In other words, the excitation pattern shifts with frequency along the cochlea without much changing its shape. The model was made *active* by having Y_b include a region of negative damping ([Neely and Kim, 1983, 1986, 2008](#)).

$$Y_b(s) = \frac{7.186 \times 10^{-2} \exp(-1.51x)}{Z_b(s)}, \quad (8)$$

$$Z_b(s) = Z_1(s) + g \times \frac{1.666 \times Z_2(s) \cdot Z_3(s)}{Z_2(s) + Z_3(s)}, \quad (9)$$

where

$$Z_1 = sM_1 + R_1 + K_1/s, \quad (10)$$

$$Z_2 = sM_2 + R_2 + K_2/s, \quad (11)$$

$$Z_3 = R_3. \quad (12)$$

$$K_1 = 1.291 \times 10^8 \exp(-3.01x), \quad (13)$$

$$R_1 = 1.797 \times 10^2 \exp(-1.51x), \quad M_1 = 0.002.$$

$$K_2 = 6.455 \times 10^1 \exp(-3.01x), \quad (14)$$

$$R_2 = 3.595 \times 10^{-1} \exp(-1.51x), \quad M_2 = 6.0 \times 10^{-9}.$$

$$R_3 = -3.593 \times 10^{-1} \exp(-1.51x). \quad (15)$$

In these equations $s = i2\pi f$ is the Laplace-transform variable, while K , R and M represent the stiffness, damping and mass, respectively. Z_1 represents the cochlear-partition impedance when the cochlea is in a passive mode, and Z_2 and Z_3 create a frequency-dependent negative-damping region. Negative damping provided cochlear amplification and a proper shape of the cochlear excitation pattern ([Neely and Kim, 1983, 1986, 2008](#)). The parameter g in Eq. (9) determined the magnitude of negative damping and the amount of cochlear amplification. When g was zero, the cochlea was in a passive mode and exhibited no amplification. The model

was piecewise linear and represented the cochlear nonlinear behavior by increasing g (producing greater amplification) with decreasing stimulus level.

The scala impedance Z_s was chosen to have a decreasing cross-sectional area to compensate for the decreasing stiffness in the cochlear-partition impedance as

$$Z_s(s) = \rho s / A_s, \quad (16)$$

$$A_s = 3.592 \times 10^{-3} \exp(-1.51x). \quad (17)$$

In this equation, $\rho=1$ represents the fluid density. These parameter values were selected to keep the characteristic impedance of the transmission line $Z_c = \sqrt{Z_s / Y_b}$ as uniform as possible to minimize reflections.

The pressure difference P has both the forward (P_f) and backward (P_r) wave components, which may be decomposed as

$$P_f = \frac{1}{2}(P - P' / \beta), \quad (18)$$

$$P_r = \frac{1}{2}(P + P' / \beta). \quad (19)$$

where $P' \equiv dP/dx$ is the x derivative (or gradient) of P , and $\beta = \sqrt{Z_s Y_b}$ is the propagation function. This decomposition formula was originally derived for a uniform transmission line (e.g., [Durney and Johnson, 1969](#)), and is exact only when the characteristic impedance is constant along the cochlea. We need to decompose the forward and backward waves only at the two ends of the cochlea, i.e., the stapes and helicotrema, for calculation of the reflectance (see below) and boundary conditions. The characteristic impedance is almost constant in these regions. Therefore, Eqs. (18) and (19) are applicable to the present model.

Cochlear reflectance, \mathfrak{R} , is defined as the ratio of the reverse-wave pressure to the forward-wave pressure wave at the stapes as

$$\mathfrak{R} = P_r / P_f. \quad (20)$$

If we assume that the effect of the middle ear is an amplitude scaling (see Sec. IV B below), then SFOAE can be modeled as the product of the cochlear reflectance, the round-trip middle ear gain, and the incident pressure at the eardrum (i.e., the stimulus) as

$$P_{\text{SFOAE}} = \mathfrak{R} \cdot G_m \cdot P_S, \quad (21)$$

where P_{SFOAE} , G_m and P_S denote the pressure of SFOAE, round-trip middle ear gain, and the pressure of stimulus in the ear canal, respectively

We first solved the transmission line equation (7) numerically for pressure P along the length of the cochlea. The cochlear length $L=35$ mm was represented by $N=3501$ points (including both end points), and the differential equation was converted into difference equations by the finite-difference method. Since we are interested in the reflection coefficient only, the boundary condition at the stapes side was set as $P=1$. At the helicotrema there should be no backward wave, which results in $P=-P'/\beta$ as the helicotrema

boundary condition. The resulting matrix equation was solved numerically for the pressure P , and the forward and backward pressures were calculated at the stapes.

B. Round-trip middle ear gain

[Puria \(2003\)](#) measured forward and reverse middle-ear pressure gain. These pressure gains varied slowly with frequency compared with SFOAE fine structures. In the current study, we are interested in the frequency range from 550 to 1450 Hz. In that range, the round-trip amplitude gain varied slowly between about -6 and -27 dB, and the phase changed between about 0.25 cycle and -0.4 cycle. From the data of [Puria \(2003\)](#), we determined that the average middle-ear round-trip amplitude gain over 500–1500 Hz was -17 dB. The phase part of the middle-ear round-trip function was negligible compared to the total phase change observed in SFOAE (i.e., NC). Therefore, we approximated the middle-ear round-trip function, G_m , with -17 dB and ignored the phase change.

C. Cochlear irregularity

Previous theoretical studies suggest that SFOAEs are generated by distributed irregularity in the cochlear-partition impedance ([Zweig and Shera, 1995](#)). It is not known which component of the impedance (i.e., stiffness, damping or mass) is most important for generation of SFOAE. Although we found that fine structures of the model cochlear reflectance could be generated with irregularity in only one of the three impedance components, irregularity in the overall impedance produced the widest range of model behavior. Therefore, we chose to apply irregularity to the overall impedance of the cochlear partition.

Irregularity in the cochlear-partition impedance was introduced by using:

$$\hat{Z}_1(x) = Z_1(x)[1 + I_r(x)], \quad (22)$$

$$I_r(x) = \alpha \cdot u_1(x) \exp[i \cdot 2\pi \cdot u_2(x)], \quad (23)$$

where $\hat{Z}_1(x)$, and $Z_1(x)$ are rough and smooth impedances, respectively, of the cochlear partition as functions of x . The relative irregularity, $I_r(x)$, is a complex quantity having a random magnitude between 0 and α ($\alpha < 1$) and a random phase between 0 and 1 cycle; $u_1(x)$ and $u_2(x)$ are uniformly distributed random numbers ranging from 0 to 1. In the initial form of the model, the irregularities at different positions along the cochlea were independent of each other.

We modified the above concept of independently distributed irregularity along the cochlear distance by introducing a novel hypothesis that the spatial-frequency content of the distributed irregularity of the cochlear-partition impedance is low-pass filtered by the cochlear-amplifier mechanism and that both the amount of cochlear amplification and the spatial filter cutoff frequency change monotonically with stimulus level. Thus, \hat{Z}_1 was represented as:

$$\hat{Z}_1 = Z_1[1 + \hat{I}_r(x)], \quad (24)$$

$$\hat{I}_r(x) = F^{-1}[F\{I_r(x)\} \cdot L(f_s)], \quad (25)$$

$$L(f_s) = \left(1 + \left|\frac{f_s}{f_{s_0}}\right|^n\right)^{-1}, \quad (26)$$

where, $F\{\cdot\}$, $F^{-1}\{\cdot\}$, f_s , f_{s_0} , and n represent spatial Fourier transform, inverse spatial Fourier transform, spatial frequency, spatial low-pass filter cutoff frequency, and the order of the filter, respectively. The spatially filtered irregularity $\hat{I}_r(x)$ can be also expressed as

$$\hat{I}_r(x) = \alpha \cdot u_3(x) \exp[i \cdot 2\pi \cdot u_4(x)], \quad (27)$$

where $u_3(x)$ and $u_4(x)$ are random numbers which have low-pass filtered spatial profiles.

D. Fit-error function

To quantify the difference between human SFOAE data (target) and the model results, we defined a fit-error function as:

$$E_{\text{fit}}(52 \text{ dB SPL}) = 0.2 \times E_{\text{NP}} + 0.2 \times E_{\text{NC}} + 0.2 \times E_{\text{EGD}} + 0.2 \times E_{\sigma_{\text{EGD}}} \quad (28)$$

$$E_{\text{fit}}(\text{other}) = 0.2 \times E_{\text{ML}} + 0.2 \times E_{\text{NP}} + 0.2 \times E_{\text{NC}} + 0.2 \times E_{\text{EGD}} + 0.2 \times E_{\sigma_{\text{EGD}}} \quad (29)$$

$$E_X = 1 - \exp[-10 \times \{(V_X - T_X)/T_X\}^2] \quad (30)$$

$$E_{\text{ML}} = 1 - \exp[-0.01 \times (V_{\text{ML}} - T_{\text{ML}})^2], \quad (31)$$

where E_X is either E_{NP} , E_{NC} , E_{EGD} or $E_{\sigma_{\text{EGD}}}$, and V_X is either V_{NP} , V_{NC} , V_{EGD} or $V_{\sigma_{\text{EGD}}}$, which represent the model values of NP, NC, EGD and STD of 25-Hz-wide EGDs, respectively; T_X represents the corresponding target value, e.g., T_{NP} corresponds to the target NP. V_{ML} is the model value of ML and T_{ML} is the target ML. The target ML at stimulus levels other than 52 dB SPL was calculated from the model ML value at 52 dB SPL with a 0.58 dB/dB slope. The 0.58 dB/dB slope corresponds to the slope of the human SFOAE ML versus stimulus level. When the model reflectance was greater than 0 dB at any frequency, the model result was rejected (the fit-error function was assigned a value of infinity).

V. MODEL RESULTS

A. Frequency-distance map of the cochlea

The frequency-distance map of the model is compared with the [Greenwood \(1990\)](#) function of the human cochlear map (Fig. 9). The two maps are in close agreement except for a region near the apical end. This study is focused on the frequency range of 550–1450 Hz where the two maps were in agreement. This agreement supports the model as a representation of the human cochlea.

B. Model SFOAE fine structures

We adjusted model parameters attempting to reproduce mean values of response features (ML, NP, NC, EGD, and

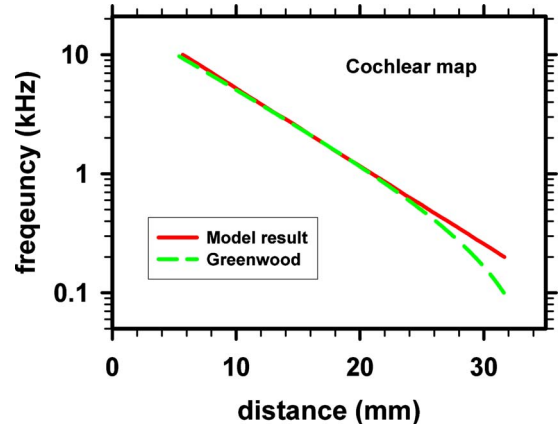


FIG. 9. (Color online) Cochlear frequency-distance map of the model (at 32 dB SPL) compared with the [Greenwood \(1990\)](#) function for the human cochlea.

σ_{EGD}) of eight subjects in the frequency range 550–1450 Hz. The above measures, except for σ_{EGD} , were plotted in Fig. 5. We generated 11 versions of the impedance irregularity spatial profiles using 11 random seeds. We obtained the mean values of the model response features among the 11 random-seed cases for each of the four stimulus levels. The four model parameters (g , α , f_{s_0} , and n) were adjusted to find a best fit for the human SFOAE data for four stimulus levels as follows. Initially we adjusted g from 0.70 to 1.00 in steps of 0.05, α from 0.02 to 0.78 in steps of 0.04, f_{s_0} from 0.50 to 1.50 cycles/mm in steps of 0.1 cycles/mm, and n from 5 to 60 in steps of 5.

During this adjustment, we first simulated the 52 dB SPL data. We treated the 52 dB data as the reference (or anchor) among the four stimulus levels. The fit-error function for 52 dB data, Eq. (28), included error components associated with four response features (NP, NC, EGD and σ_{EGD}) but excluded an error component associated with ML. We chose this approach because the model exhibited a difficulty of fitting all features including ML. Once we obtained the optimal model parameters for a best fit of the 52 dB data based on Eq. (28), we then attempted to simulate the optimal model behavior for other stimulus levels using all features including ML, i.e., using Eq. (29). This approach was intended to reproduce an optimal slope of SFOAE-level versus stimulus level. Accordingly, we derived target values of ML for 32, 42 and 62 dB SPL by combining the model ML at 52 dB and a slope of 0.58 dB/dB. The latter slope was obtained as the slope of a linear regression line fit for the human data of ML versus stimulus level [Fig. 5(A)].

During the search for optimal model parameters (g , α , f_{s_0} , and n), we imposed the constraint that each of the model parameters should vary monotonically and smoothly with stimulus level. A smooth change meant that the slope of each parameter versus stimulus level should also vary monotonically with stimulus level.

We defined the total fit error to be the sum of four fit errors for the four stimulus levels. After finding the initial candidate optimal parameter set that minimized the total fit error based on the coarse steps of the parameter values, we performed a further search with fine steps of each parameter

TABLE V. Means (\bar{m}) and standard deviations (σ) of six response features of model SFOAE results obtained with 11 random seeds of the cochlear-partition irregularity under the condition of the optimal values of the model parameters. The response features were: mean level (ML) of SFOAE in dB SPL; number of peaks (NP); frequency spacing (Δf) between SFOAE peaks in Hz; number of cycles (NC) of SFOAE phase change; energy-weighted group delay (EGD) in ms; and standard deviation of EGD (σ_{EGD}) in ms.

| Stimulus level (dB SPL) | | Model result | | | | | |
|-------------------------|-----------|--------------|------|------------|-----|-----|-----------------------|
| | | ML | NP | Δf | NC | EGD | σ_{EGD} |
| 32 | \bar{m} | -13.9 | 15.0 | 56.6 | 7.2 | 9.0 | 6.8 |
| | σ | 1.5 | 1.2 | 4.3 | 2.3 | 2.7 | 1.8 |
| 42 | \bar{m} | 7.0 | 11.1 | 77.8 | 7.7 | 8.1 | 5.4 |
| | σ | 1.1 | 1.4 | 12.2 | 1.3 | 1.0 | 1.3 |
| 52 | \bar{m} | 1.5 | 7.1 | 111.7 | 7.1 | 7.9 | 3.8 |
| | σ | 1.6 | 1.0 | 24.8 | 0.8 | 1.5 | 0.7 |
| 62 | \bar{m} | 4.0 | 5.5 | 151.8 | 5.3 | 6.1 | 3.8 |
| | σ | 1.7 | 0.9 | 20.2 | 1.3 | 1.1 | 0.8 |

around the initial candidate optimal parameter values. The fine step size was 0.01 for g , 0.02 for α , 0.5 cycles/mm for f_{s_0} , and 1 for n . The set of model parameters that yielded the smallest average total fit error across the 11 random-seed cases representing the response features ML, NP, NC, EGD and σ_{EGD} , was chosen as *the optimal parameter set*. The model response features obtained with the optimal parameter set are summarized in Table V, and the optimal parameters of the model and the fit errors are shown in Table VI.

Figures 10 and 11 show simulated SFOAE level, phase, and 25-Hz-wide EGD versus frequency obtained with the optimal parameter set of the model and with two particular random seeds Nos. 6 and 1, respectively. The model exhibited SFOAE characteristics similar to those of the human SFOAEs shown in Figs. 2 and 3. Thus, the model successfully reproduced salient features of the human SFOAEs at stimulus levels of 32–62 dB SPL.

Analogous to the human SFOAEs, major troughs of SFOAE level of the model also tended to have concomitant rapid changes in phase slope and EGD values. For example, the SFOAE level in Fig. 10(G) exhibited major troughs at ~ 720 and 980 Hz. Around these frequencies, there were concomitant rapid changes in the phase slopes [Fig. 10(H)] and EGDs [Fig. 10(I)]. Analogous examples of troughs [Fig. 11(A)] and rapid changes of phase slopes [Fig. 11(B)] and

EGDs [Fig. 11(C)] are also visible around 810 and 1210 Hz in results obtained with a different random seed.

The confidence intervals of model parameters were determined as follows. Within the confidence interval of a parameter, the fit error did not exceed twice the optimal fit error among the 11 random-seed cases. That is, the model generated sufficiently similar features of SFOAEs when the model parameters were within the confidence intervals. The confidence intervals are shown in Table VII. For example, for 52 dB SPL, optimal fit-error E_{fit} was 0.040 among 11 cases (Table VI). If we vary g in a range of 0.87–0.91 while fixing the other three parameters, then the fit error remained smaller than $2 \times E_{\text{fit}}$, i.e., 0.080. In some cases, fit-error values were

TABLE VI. Optimal values of the model parameters that produced minimal fit-errors (E_{fit}). E_{fit} , dimensionless, is defined in Eqs. (28) and (29). The model parameters were g (dimensionless amplification parameter), α (dimensionless scale factor for the cochlear-partition impedance irregularity), f_{s_0} (cutoff frequency, in cycles/mm, of a low-pass spatial-frequency filter acting on the irregularity), and n (dimensionless parameter for the slope of the low-pass spatial-frequency filter).

| Stimulus level (dB SPL) | g | α | f_{s_0} | n | E_{fit} |
|-------------------------|------|----------|-----------|-----|------------------|
| 32 | 0.93 | 0.42 | 0.85 | 9 | 0.13 |
| 42 | 0.91 | 0.22 | 1.20 | 13 | 0.073 |
| 52 | 0.88 | 0.10 | 1.40 | 24 | 0.040 |
| 62 | 0.73 | 0.08 | 1.40 | 50 | 0.053 |

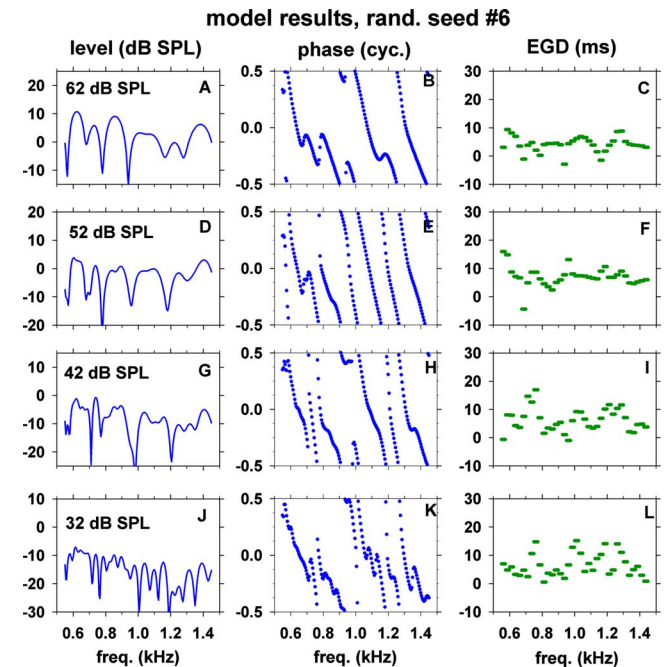


FIG. 10. (Color online) SFOAE fine structures of the model obtained with random seed No. 6 for the cochlear-partition impedance irregularity. The three columns correspond to the model SFOAE level, phase and energy-weighted group delay (EGD) versus frequency. The four rows correspond to four stimulus levels indicated in the insets. The model parameters were set to the optimal values described in Table VI.

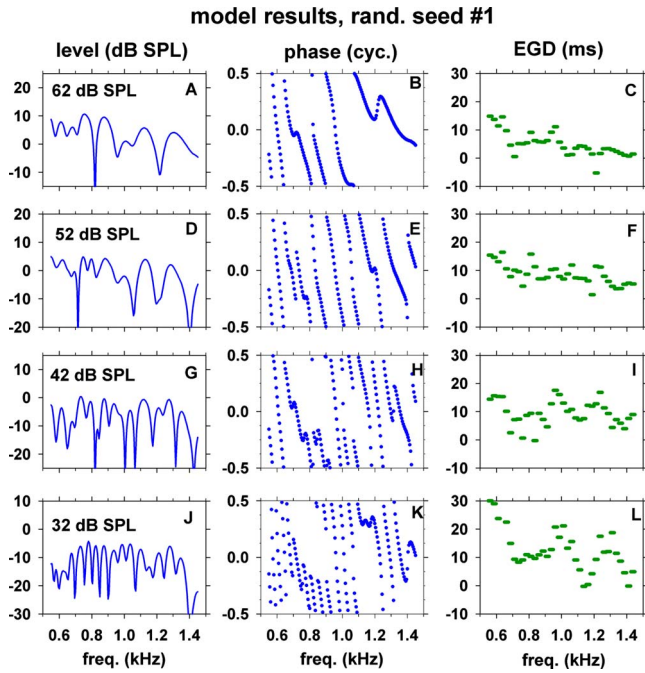


FIG. 11. (Color online) SFOAE fine structures of the model. All conditions of the model were the same as those of Fig. 10 except for the use of random seed No. 1 for the cochlear-partition impedance irregularity.

quite sensitive to the parameter change so that we introduced very small value ε which is smaller than 0.025 for f_{s_0} and 0.5 for n to represent the confidence interval. However, n value for 62 dB SPL in Table VII could be increased indefinitely because there was little effect on the low-pass filter when $n > 16$. Generally, the confidence intervals of the four model parameters shown in Table VII were narrow. That is, the model parameters for different stimulus levels were distinct with little overlaps. This indicates that the model parameters were well constrained by the human SFOAE features.

Changes of the model response features as results of changes in the model parameters are described in Table VIII using the 52 dB SPL stimulus condition as an example. Table VIII includes the optimal parameter set ($g=0.88$, $\alpha=0.10$, $f_{s_0}=1.40$ cycles/mm and $n=24$) together with other points of the parameter space corresponding to the extremes of the confidence interval of each parameter. Values of the model response features for the above conditions are shown. Within the confidence intervals, the relative magnitudes of changes

in ML, NP, Δf , NC, EGD and Q_3 dB were less than 2.2 dB, 8.5%, 9.7%, 8.5%, 11.4% and 4.2%, respectively. Within the intervals, increases of g , α , or f_{s_0} led to increases of ML, NP, NC and EGD. A decrease of n led to a similar effect. Regarding Q_3 dB, only g among the four parameters affected Q_3 dB such that an increase of g led to an increase of Q_3 dB.

C. Spatial profiles of irregularity

Spatial profiles of the amplitude (left column) and phase (right column) of the impedance irregularity under the condition of the optimal model parameters are described in Fig. 12. At 32 dB SPL, where the spatial low-pass cutoff frequency was the lowest, the spatial profile of amplitude and phase of the irregularity varied most slowly over cochlear distance among the four cases corresponding to four stimulus levels. With increasing stimulus level, the spatial low-pass cutoff gradually increased leading to increasingly more rapid changes of irregularity over distance.

D. Estimated human cochlear excitation patterns and transfer functions

Using the optimal model parameters described above, we determined cochlear excitation patterns and transfer functions of the model cochlear partition. The excitation patterns for 1 kHz stimulus frequency, corresponding to the ratio of cochlear-partition velocity to stapes velocity as a function of cochlear distance for a fixed frequency, are shown in Fig. 13, panels A and B. With decreasing stimulus level, the excitation pattern became sharper and the traveling-wave peak became higher moving slightly apically. The total phase lag of the excitation pattern was more than 15 cycles over the full cochlear distance. At the highest stimulus level, the phase-versus-distance slope became shallower.

The cochlear-partition transfer functions, corresponding to the ratio of cochlear-partition velocity to stapes velocity as a function of stimulus frequency for a fixed cochlear distance (i.e., the 1 kHz place), are shown in Fig. 13, panels C and D. The shapes of the partition responses versus distance (panel C) were quite similar to the partition responses versus stimulus frequency (panel A).

The sharpness of the partition transfer functions were quantified in terms of three types of quality factors and peak-to-shoulder ratio (PSR), also known as the tip-to-tail ratio.

TABLE VII. Confidence intervals of the model parameters g , α , f_{s_0} , and n . The confidence interval of each parameter was defined as one where the average fit error among the 11 random seed cases remained smaller than twice the optimal fit error. The confidence interval was determined by varying only one parameter at a time with the other parameters fixed at the optimal value (described in Table VI). The units of f_{s_0} are cycles/mm and the other three parameters are dimensionless. The symbol, ε , is intended to indicate a small quantity; it was less than the following values: 0.025 for f_{s_0} , and 0.5 for n .

| Parameters | Stimulus level (dB SPL) | | | |
|------------|---|---|-----------|---|
| | 32 | 42 | 52 | 62 |
| g | 0.91–0.94 | 0.90–0.92 | 0.87–0.91 | 0.67–0.81 |
| α | 0.16–0.64 | 0.18–0.26 | 0.08–0.12 | 0.06–0.16 |
| f_{s_0} | $(0.85-\varepsilon)-(0.85+\varepsilon)$ | $(1.20-\varepsilon)-(1.20+\varepsilon)$ | 1.40–1.45 | $(1.40-\varepsilon)-(1.40+\varepsilon)$ |
| n | $(9-\varepsilon)-(9+\varepsilon)$ | $(13-\varepsilon)-(13+\varepsilon)$ | 20–28 | 16– ∞ |

TABLE VIII. Changes in model response features as results of changes in each model parameter at 52 dB SPL stimulus level. The range of each parameter was the confidence interval of the parameter described in Table VII. See the caption of Table VI for definitions of the model parameters.

| Parameter | Model response features | | | | | | | |
|--------------------|-------------------------|------|--------------------|--------------|-------------|-------------------------------|--------------------|-----|
| | ML (dB SPL) | NP | Δf (Hz) | NC (cyc.) | EGD (ms) | σ_{EGD} (ms) | $Q_{3 \text{ dB}}$ | |
| Optimal parameters | -1.5 | 7.1 | 111.7 | 7.1 | 7.9 | 3.8 | 7.1 | |
| g | 0.87 | -1.8 | 6.9 | 115.5 | 7.0 | 7.7 | 3.8 | 7.0 |
| | 0.91 | -0.8 | 7.4 | 106.3 | 7.5 | 8.5 | 4.2 | 7.4 |
| α | 0.08 | -3.5 | 6.5 | 122.5 | 7.0 | 7.7 | 3.7 | 7.1 |
| | 0.12 | 0.2 | 7.1 | 112.5 | 7.4 | 8.1 | 4.1 | 7.1 |
| f_{s_0} | 1.40 | -1.5 | 7.1 | 111.7 | 7.1 | 7.9 | 3.8 | 7.1 |
| | 1.45 | 0.7 | 7.4 | 106.9 | 7.7 | 8.8 | 3.9 | 7.1 |
| n | 20 | -1.0 | 7.3 | 108.6 | 7.2 | 8.2 | 4.1 | 7.1 |
| | 28 | -1.7 | 6.6 | 117.9 | 6.9 | 7.8 | 3.9 | 7.1 |

Quality factors $Q_{3 \text{ dB}}$ and $Q_{10 \text{ dB}}$ correspond to the center frequency divided by the width of transfer function measured 3 and 10 dB below the peak, respectively. Q_{ERB} correspond to an analogous measure where the width of an equivalent rectangular bandwidth is used. For PSR, the shoulder corresponds to the part of the transfer function at a frequency below the peak where the curves of different stimulus levels converge. The values of these parameters for the model partition transfer functions are indicated in Table IX. $Q_{3 \text{ dB}}$ and Q_{ERB} ranged 5.5–8.0, $Q_{10 \text{ dB}}$ 3.2–4.3, and PSR 38–60 dB.

E. Decomposition of model SFOAE into long- and short-delay components

To determine the cochlear locations where SFOAEs are generated in the model, we examined the model behavior

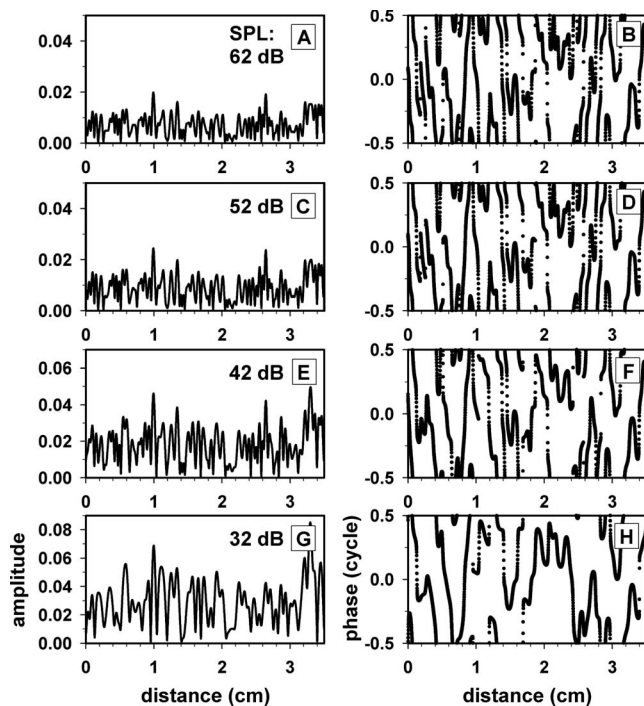


FIG. 12. Spatial profiles of amplitude (left column) and phase (right column) of the model cochlear-partition impedance irregularity. The four rows correspond to four stimulus levels. The model parameters were set to the optimal values described in Table VI.

with the impedance irregularity confined either in a region around the traveling-wave peak or in regions remote from the peak. The traveling-wave peak region was associated with steeper phase-versus-distance slope [Figs. 13(B) and 14(B)]. In terms of wavelength (λ), i.e., the cochlear distance spanned by one cycle of phase change, the peak region has shorter λ than the remote regions [Fig. 14(C)]. We defined the peak region to be a cochlear region with $\lambda < 2.5$ mm, and the remote regions to be the remaining cochlear regions.

Besides the normal impedance irregularity distributed throughout the full cochlear length, we also prepared two other types of impedance irregularity profile: (1) irregularity restricted to the peak region with $\lambda < 2.5$ mm, and (2) irregularity restricted to the remote regions with $\lambda > 2.5$ mm.

The model SFOAEs obtained with three types of irregu-

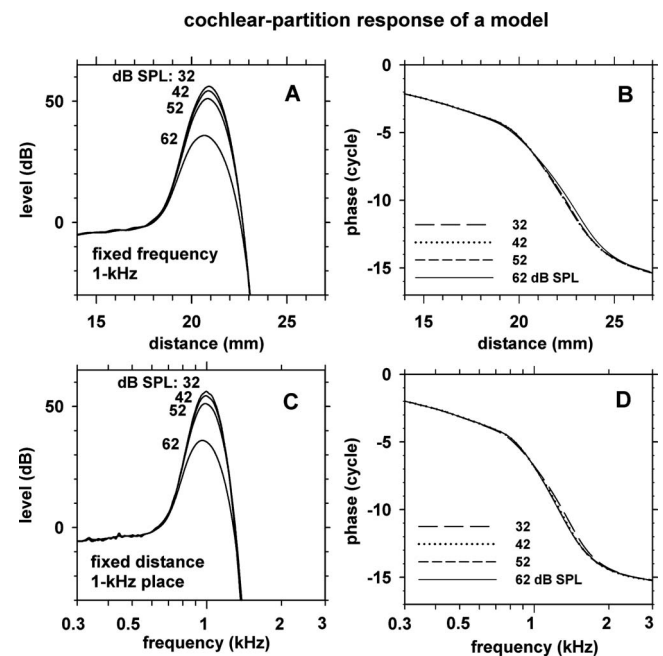


FIG. 13. Cochlear-partition response level and phase versus cochlear distance for 1 kHz stimulus frequency (panels A and B) and versus stimulus frequency for the 1 kHz cochlear place (panels C and D). Four curves in each panel correspond to four stimulus levels (32–62 dB SPL) indicated in the insets. The cochlear-partition response was relative to the stapes input. The model parameters were set to the optimal values described in Table VI.

TABLE IX. Response characteristics of the model cochlear-partition transfer functions at the 1 kHz place at stimulus levels of 32–62 dB SPL (described in Fig. 13). Means and standard deviations (STDs) were calculated for 11 random-seed cases. The results were obtained with the optimal model parameters described in Table VI. $Q_{3\text{ dB}}$ and $Q_{10\text{ dB}}$ are “quality factors” measured at 3 and 10 dB points relative to the peak; Q_{ERB} is a quality factor measured using the equivalent rectangular bandwidth (ERB). The quality factors are dimensionless. PSR stands for peak-to-shoulder ratio, also known as tip-to-tail ratio.

| | Stimulus Level (dB SPL) | | | | | | | |
|--------------------|-------------------------|------|------|------|------|------|------|------|
| | 32 | | 42 | | 52 | | 62 | |
| | Mean | STD | Mean | STD | Mean | STD | Mean | STD |
| $Q_{3\text{ dB}}$ | 8.0 | 0.53 | 7.6 | 0.30 | 7.1 | 0.19 | 5.8 | 0.07 |
| Q_{ERB} | 7.6 | 0.58 | 7.1 | 0.24 | 6.6 | 0.12 | 5.5 | 0.06 |
| $Q_{10\text{ dB}}$ | 4.3 | 0.24 | 4.1 | 0.12 | 3.9 | 0.05 | 3.2 | 0.02 |
| PSR (dB) | 60 | 1.6 | 57 | 0.8 | 54 | 0.3 | 38 | 0.2 |

larity profiles at 32 dB SPL are shown in Fig. 15. The three rows correspond to the three types of irregularity profiles: fully present (top), restricted to the peak region (middle), or to the remote regions (bottom). As the spatially restricted irregularity profiles were specifically associated with the 1 kHz traveling-wave peak, the model results of Fig. 15 are shown only for a narrow range of frequencies, 950–1050 Hz, around 1 kHz. The phase slope of the model SFOAE generated by the irregularity in the peak region [Fig. 15(D)] was steep with an EGD of 20.8 ms. The ratio of 20.8 ms to EGD of the 1 kHz place of the cochlea at 32 dB SPL (12.6 ms) is 1.7. This is slightly smaller than 2, expected from a situation where the signal propagates to the traveling-wave peak and back to the ear canal. In contrast, the phase slope of SFOAE generated by the remote regions [Fig. 15(F)] was shallow with an EGD of 4.1 ms, which is much shorter than the EGD of the 1 kHz cochlear place (12.6 ms.).

The results described in Fig. 15 demonstrate that, in the model, (1) the SFOAE consists of a long-delay component generated by irregularity in the traveling-wave peak region and a short-delay component generated by irregularity in the remote regions, (2) the levels of the long- and short-delay components are comparable to each other and they both change slowly with frequency, and (3) the rapidly changing fine structure of SFOAE of the model is a result of quasi-periodic vector cancellations between the long- and short-delay components.

VI. DISCUSSION

A. Main results

The main results of the present study are: (1) introduction of a novel method of efficiently measuring SFOAE level and phase as functions of frequency by a combined use of frequency-swept stimuli and the digital heterodyne analysis; (2) measurement of high-resolution data of SFOAE level, phase, and EGD versus frequency in normally hearing human adult subjects; (3) reproduction of salient features of the human SFOAEs in an active cochlear model; (4) inferences about characteristics of human cochlear-partition motion, and (5) demonstration that, in the model, an SFOAE consists of a long-delay component generated by irregularity in the traveling-wave peak region and a short-delay component generated by irregularity in cochlear regions remote from the peak.

B. Novelty and advantages of the present method of SFOAE measurement

The novelty of the present method lies in an unprecedented efficiency of measuring high-resolution SFOAEs by a combined use of *continuously* sweeping the stimulus frequency and the digital heterodyne analysis. The frequency resolution of the present method depends on the square root of the frequency sweep rate as stated in Methods, Eq. (3). Accordingly, a twofold higher frequency resolution can be

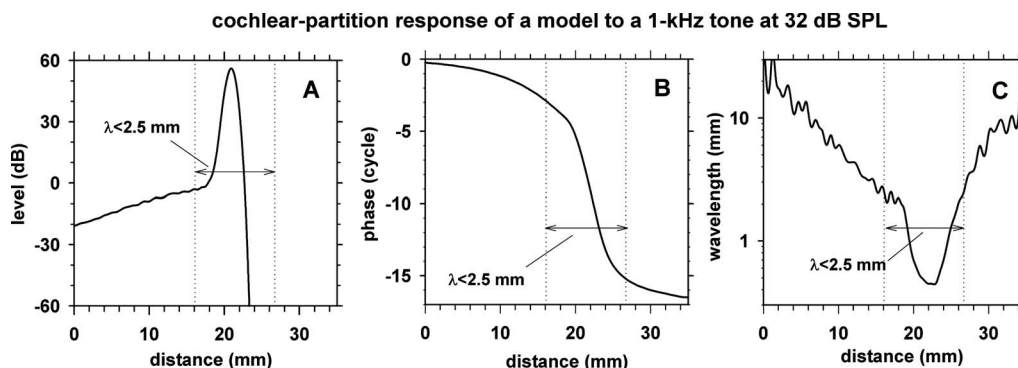


FIG. 14. Cochlear-partition response to a 1 kHz 32 dB SPL stimulus in terms of level (A), phase (B) and wavelength (C) as functions of cochlear distance. The cochlear-partition response was relative to the stapes input. The model parameters were set to the optimal values described in Table VI.

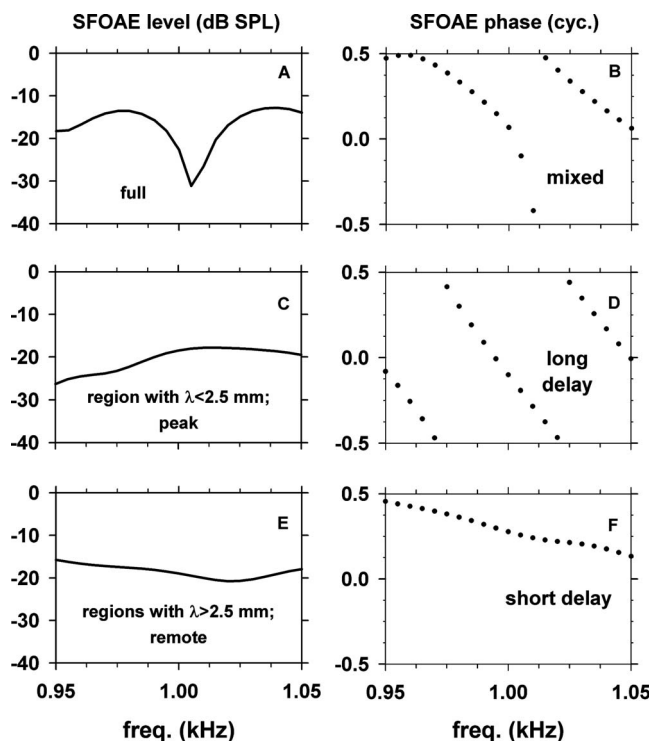


FIG. 15. Level and phase of the model SFOAE at 32 dB SPL versus frequency over a narrow range of frequencies (950–1050 Hz) obtained with three types of spatial profiles of the cochlear-partition impedance irregularity: fully present (*top*), restricted in a region around the traveling-wave peak with $\lambda < 2.5$ mm (*middle*) and in regions remote from the peak with $\lambda > 2.5$ mm (*bottom*).

achieved by decreasing the sweep rate four times. Equivalently, a fourfold faster sweep rate can be achieved if one is willing to decrease frequency resolution by a factor of 2.

In the conventional methods of SFOAE measurement (Shera and Guinan, 1999; Shera and Guinan, 2003; Goodman *et al.*, 2003; Schairer *et al.*, 2006), SFOAE fine structures are measured with many probe frequencies, one frequency at a time. Consequently, the conventional method has a greater overhead, i.e., rise and fall gating times and the times involved for saving the data on a storage device for each individual frequency rendering it inefficient.

An efficient recording method facilitates data collection with less variability and a higher signal-to-noise ratio for a given recording time. Data obtained over long recording times are more susceptible to artifacts arising from unavoidable random movements of the subjects making the data more variable.

C. Future applications of swept stimuli and heterodyne analysis

The present method can be used to obtain SFOAEs over a wider range of frequencies, e.g., 500–8000 Hz (Hill *et al.*, 2008). In such a study, an exponential frequency sweep is more desirable than a linear frequency sweep. A simple variation of the present method will allow one to measure DPOAEs as a continuous function of either primary frequencies (Long *et al.*, 2008) or primary levels. For example, one may sweep f_1 , f_2 or f_2/f_1 over a period of time. The heterodyne analysis will then yield DPOAE level and phase (e.g.,

at $2f_1 - f_2$) for various f_2 , for example, by extracting DPOAE level and phase at various time points. A similar concept of measuring input-output functions of DPOAE level was introduced by Neely *et al.* (2003).

D. Why is SFOAE group delay far shorter than expected?

Prevailing concepts about SFOAEs are that: (1) SFOAEs are generated predominantly in a cochlear region around the traveling-wave peak, and (2) the SFOAE group delay at a frequency are approximately twice the group delay of the cochlear partition at the traveling-wave peak (Zweig and Shera, 1995). Siegel *et al.* (2005) observed that the group delays of SFOAEs in chinchillas were similar to (for frequencies >4 kHz), or shorter than (for frequencies <4 kHz) the group delays of the cochlear partition at the characteristic frequencies (the partition group delays were estimated from responses of cochlear nerve fibers in chinchillas). This observation is *inconsistent* with the prevailing concept indicated above. Siegel *et al.* highlighted this discrepancy as an important issue for the hearing-science field.

Shera *et al.* (2006) suggested that one approach to resolve the discrepancy between the Zweig–Shera model prediction and the Siegel *et al.* observation is to hypothesize that the total SFOAE is a mixture of a reflection-source emission component with a long delay and a nonlinear-distortion-source emission component with a short delay. (A similar suggestion was made by Goodman *et al.*, 2003) After Shera *et al.* separated (“un-mixed”) the total SFOAE into short- and long-delay components, they found that the group delays of the long-delay component of SFOAE were closer to the predicted group delays. However, the group delays of the long-delay SFOAE component were still shorter than the predicted group delays for all frequencies with the differences being greater for frequencies <4 kHz. Furthermore, it remains unknown whether the short-delay component of SFOAE indeed represents a distortion-source emission or some other type of reflection-source emission.

The results of the present cochlear model (Fig. 15) provide a novel hypothesis that the short- and long-delay components of SFOAE are both reflection-source emissions but arise from different regions of the cochlea. In the model, the short-delay SFOAE component arises from irregularity located in cochlear regions remote from the traveling-wave peak whereas the long-delay component arises from irregularity in the peak region. Because the model is piecewise linear, there is no nonlinear-distortion-source emission in the model. Future experiments should verify the validity of this hypothesis.

E. Comparison with the coherent reflection model

The coherent reflection model of SFOAE [Zweig and Shera, 1995, Eq. (23), page 2023] states that reflected wavelets of a forward cochlear traveling wave combine in phase (i.e., creates coherent reflection) when $2\Delta x = \lambda$, or $\Delta x = \lambda/2$, where λ is the wavelength and Δx is distance between cochlear mechanical irregularities. An equivalent statement

is that reflected wavelets create coherent reflection when the spatial-frequency (f_s) content of the irregularity profile is large at $f_s=2/\lambda$.

In our model, λ at the peak of the traveling-wave envelope is ~ 0.5 mm for a 1 kHz sinusoidal stimulus at 32 dB SPL [Fig. 14(C)]. Thus, in order to produce coherent reflection at the traveling-wave peak, the spatial-frequency content of the irregularity profile has to be large at $f_s=2/\lambda=2/0.5=4$ cycles/mm.

A region of the cochlear model that includes the traveling-wave peak and nearby regions has shorter wavelengths [$\lambda=0.45\text{--}2.5$ mm, Fig. 14(C)] than the basal region further away from the peak [$\lambda=2.5\text{--}25$ mm, Fig. 14(C)]. In the peak region with $\lambda=0.45\text{--}2.5$ mm, generation of coherent reflection requires that the spatial-frequency content of the irregularity profile has to be large at $f_s=2/\lambda=0.8\text{--}4$ cycles/mm.

In contrast, the irregularity profiles in the model were filtered with a steep low-pass spatial-frequency filter with a cutoff frequency of 0.85 cycles/mm at 32 dB SPL; this parameter was obtained as a result of an objective fitting procedure. Therefore, the spatial-frequency content of the model irregularity profile was quite small at the 0.8–4 cycles/mm required for generation of coherent reflection in the short- λ peak region. Consequently, the SFOAE originating from the peak region of the model was much lower than expected from the coherent-reflection model which assumes that the spatial-frequency content of the irregularity profile is large at $f_s=2/\lambda$.

Because λ is much longer (2.5–25 mm) in the basal region, the spatial frequency for large irregularity required by coherent reflection is much lower ($f_s=2/\lambda=0.08\text{--}0.8$ cycles/mm) in the basal region than in the peak region. The low-pass spatial-frequency filter with a cutoff frequency of 0.85 cycles/mm preserves amplitudes of irregularity at the above required spatial frequencies. Therefore, the irregularity profile in the long- λ basal region of the model was in a more favorable condition for generation of coherent reflection than the irregularity profile in the short- λ peak region.

The above considerations provide an explanation of a surprising behavior of the present cochlear model. That is, in the model, the levels of the short- and long-delay SFOAE components originating from the remote and peak regions, respectively, are comparable to each other, thus creating rapidly changing peaks and troughs of SFOAE level versus frequency.

F. Comparison with previous studies

Changes in SFOAE fine structures with stimulus levels were reported by Goodman *et al.* (2003). Stimulus-level-dependent changes of frequency spacing, observed in the Goodman *et al.* study and the present study, represent a nonlinear behavior. This behavior was simulated by the present model with stimulus-level-dependent changes of the model parameters (g , α , f_{s_0} , and n).

Our observation of an increase of normalized group delay with frequency over 500–1500 Hz is consistent with pre-

vious studies (Shera and Guinan, 2003; Schairer *et al.*, 2006). The present observation of a decreasing group delay of SFOAE with increasing stimulus level (Figs. 4 and 5) is consistent with Schairer *et al.*, (2006). Mechanical responses of a basal part of the chinchilla cochlea also exhibited an analogous decrease of group delay with increasing stimulus level (Ruggero *et al.*, 1997).

G. Model assumptions

Although the model results successfully reproduced the six features of the human SFOAE fine structures (e.g., Fig. 5 and Table V), some differences between the model results and the human data were noticeable. One such example is the SFOAE phase versus frequency pattern. The human phase-frequency pattern (Figs. 2 and 3) tended to be more evenly spread out across the 550–1450 Hz range than the model counterpart (Figs. 10 and 11). It is not clear what changes in the model can improve in this regard.

There were other shortcomings of the model. We used a piecewise-linear frequency-domain model of the cochlea. Development of a nonlinear time-domain version of the model will be a desirable future goal. We used a one-dimensional cochlear model even though a cochlear model should ultimately be three dimensional. We believe, however, that important insights were gained by using a simplified one-dimensional model. Exploration of the parameter space of the model was facilitated by the simplified form of the model.

In addition, the boundary condition at the stapes did not include properties of the middle ear. A lossy-impedance boundary condition provided by the middle ear at the stapes should better represent reflection of the backward wave at the stapes (e.g., Shera and Zweig, 1991; Zweig and Shera, 1995; Talmadge *et al.*, 1998, 2000). Since the round-trip reflectance, i.e., the multiplication of reflectance from the cochlear-partition irregularities and the reflectance from the stapes, is small, the effects of multiple reflections may be small for the cases presented in this paper.

H. Is spatial profile of irregularity controlled by the cochlear amplifier?

In the model, the model parameters (g , α , f_{s_0} , and n), co-varied with stimulus level (Table VI). Specifically, as the stimulus level decreased, the model became more active (amplification increased), the base amplitude of the irregularity increased, and the spatial low-pass filter exhibited a lower cutoff and a steeper filter slope. These hypothesized relationships among the model parameters and stimulus level were found as a result of the objective fitting procedure searching for the optimal model parameters that best reproduced the human SFOAE fine structures at various stimulus levels.

A possible implication of the present hypothesis is that the spatial frequency content of the cochlear-partition impedance irregularity profile is controlled by the cochlear amplifier mechanism. When the model becomes more active at a lower stimulus level, the hypothesis implies that the active

mechanism exerts a more prominent longitudinal coupling of the organ of Corti whereby the partition impedance changes more slowly along the cochlear distance.

I. Inferences about the human cochlea

The present human SFOAE data together with an active cochlear model allowed us to make inferences about characteristics of the human cochlear-partition responses. The detailed and reliable measurements of SFOAE fine structures proved valuable in guiding the process of adjusting parameters of an active cochlear model. The optimal model parameters were well constrained by the six features of the SFOAE fine structures (Table V) with narrow confidence intervals (Tables VII and VIII). The present model suggests that: (1) the human cochlear-partition response at the 1 kHz place at 32 dB SPL has a sharp tuning ($Q_{3\text{ dB}}$ of 8.0) and a large amplification (peak-to-shoulder ratio of 60 dB), and (2) the tuning becomes broader and the amplification smaller as the stimulus level increased to 62 dB SPL.

Psychophysically observed Q_{ERB} of the human cochlear tuning at 1 kHz at a low stimulus level have been reported to be 7.5 (Glassberg and Moore, 1990), 12.7 (Shera *et al.*, 2002), and 11 (Oxenham and Shera, 2003). The present cochlear-model prediction of $Q_{\text{ERB}}=7.6$ at 32 dB SPL is close to the value of Glassberg and Moore.

To compare the frequency-tuning characteristics of the present model with those of animal cochlear nerve fibers, we used the facts that the human 1 kHz cochlear place is located 60% from the stapes (Figs. 9 and 13) and that a 60% position corresponds to 2.8 kHz in cat (Liberman, 1982), 2 kHz in guinea pig (Wilson and Johnstone, 1975) and 1 kHz in chinchilla (Eldredge *et al.*, 1981). Because the tuning sharpness of a cochlear nerve fiber changes with the fiber's characteristic frequency (CF), it is appropriate to use a common relative cochlear position (60% from the stapes in this case) in a comparison of tuning sharpness across species. In cat, threshold tuning curves of cochlear nerve fibers with CF = 2.8 kHz show an average $Q_{10\text{ dB}}=4.5$ (Liberman, 1990). In guinea pig, cochlear nerve fibers and medial olivocochlear nerve fibers with CF \approx 2 kHz show $Q_{10\text{ dB}} \approx 2.3 \sim 4$ (Pickles, 1984; Brown *et al.*, 1998). In chinchilla, cochlear nerve fibers with CF \approx 1 kHz show $Q_{10\text{ dB}} \approx 1.7$ (Temchin *et al.*, 2005). Thus, the cochlear places at 60% from the stapes in these species exhibit a diverse range of $Q_{10\text{ dB}}$. The present human model prediction of $Q_{10\text{ dB}}=4.3$ at 32 dB SPL is similar to the $Q_{10\text{ dB}}$ value of cat (Liberman, 1990).

The present model predictions about the human cochlea are consistent with cochlear-partition responses of animals regarding the shapes of the amplitude and phase parts of the transfer function and stimulus-level-dependent changes in several response measures: (1) broadening of frequency tuning with increasing stimulus level, and (2) decrease of peak-to-shoulder ratio with increasing stimulus level (Rhode, 1971, 2007; Sellick *et al.*, 1982; Ruggero *et al.*, 1997; Robles and Ruggero, 2001; Ren *et al.*, 2006). Peak-to-shoulder ratio is related to the cochlear-amplifier gain. Cochlear amplifier gain, defined in various ways, has been observed in animals to be in a range of 40–80 dB (Ruggero *et al.*, 1997; Rhode,

2007). Peak-to-shoulder ratio of basal points of the chinchilla cochlear partition (CF=5 ~ 10 kHz) were observed to be 33–50 dB at low stimulus levels (Ruggero *et al.*, 1997; Rhode, 2007). The present prediction of peak-to-shoulder ratio of the human cochlea, 60 dB at a stimulus level of 32 dB SPL, is slightly higher than those observed in the chinchilla cochlear partition.

ACKNOWLEDGMENTS

This study was supported in part by USA NIH-NIDCD Grant Nos. R01DC00360 (PI, D.O.K.) and R01DC08318 (PI, S.T.N.). The experiments were conducted while Y.-S.C. was visiting the University of Connecticut Health Center supported by Brain Korea 21 Project, the School of Information Technology, KAIST, in 2002–2003. Y.-S.C. and S.-Y.L. have been supported partly by the Brain Neuroinformatics Research Program sponsored by Korean MOST and MOCIE from 2001. We thank Dr. Gerhard Hill for assistance in making measurements using the conventional method, Dr. S. Puria for providing numerical middle-ear data, Dr. M. Ruggero and Dr. A. Temchin for providing numerical cochlear-partition data. We also thank Dr. J. Siegel, Dr. C. Shera, and Dr. M. Ruggero for helpful discussions on topics of this study.

- American National Standards Institute (1978). Methods for Manual Pure-Tone Threshold Audiometry. ANSI S3.21-1978 (reaffirmed 1986).
- Brass, D., and Kemp, D. T. (1993). "Suppression of stimulus frequency otoacoustic emissions," *J. Acoust. Soc. Am.* **93**, 920–939.
- Brown, M. C., Kujawa, S. G., and Liberman, M. C. (1998). "Single Olivocochlear Neurons in the guinea pig. II. Response plasticity due to noise conditioning," *J. Neurophysiol.* **79**, 3088–3097.
- Durney, C. H., and Johnson, C. C. (1969). *Introduction to Modern Electromagnetics* (McGraw-Hill, New York), pp. 365–368.
- Eldredge, D. H., Miller, J. D., and Bohne, B. A. (1981). "A frequency-position map for the chinchilla cochlea," *J. Acoust. Soc. Am.* **69**, 1091–1095.
- Glassberg, B. R., and Moore, B. C. J. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.
- Goldstein, J. L., Baer, T., and Kiang, N. Y. S. (1971). "A theoretical treatment of latency, group delay and tuning. Characteristics for auditory nerve responses to clicks and tones," in *The Physiology of the Auditory System*, edited by M. B. Sachs (National Educational Consultants, Baltimore), pp. 133–141.
- Goodman, S. S., Withnell, R. H., and Shera, C. A. (2003). "The origin of SFOAE microstructure in the guinea pig," *Hear. Res.* **183**, 7–17.
- Greenwood, D. D. (1990). "A cochlear frequency-position function for several species—29 years later," *J. Acoust. Soc. Am.* **87**, 2592–2605.
- Guinan, Jr., J. J., Backus, B. C., Lilaonitkul, W., and Aharonson, V. (2003). "Medial olivocochlear efferent reflex in humans: Otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency OAEs," *J. Assoc. Res. Otolaryngol.* **4**, 521–540.
- Hill, G., Parham, K., Choi, Y.-S., and Kim, D. O. (2008). "Stimulus-frequency otoacoustic emissions in normally-hearing humans obtained with a frequency-sweeping method over 0.5 to 8 kHz," *Assoc. Res. Otolaryngol. Meeting*, Abstr. No. 186.
- Kemp, D. T. (1978). "Stimulated acoustic emissions from within the human auditory system," *J. Acoust. Soc. Am.* **64**, 1386–1391.
- Kemp, D. T. (1986). "Otoacoustic emissions, traveling waves and cochlear mechanisms," *Hear. Res.* **22**, 95–104.
- Kemp, D. T., and Chum, R. A. (1980). "Observations on the generator mechanism of stimulus frequency acoustic emissions—Two tone suppression," in *Physiological, Psychological, and Behavioral Studies in Hearing*, edited by G. van der Brink and F. A. Bilten (Delft University Press, Delft, Netherlands), pp. 34–42.
- Kemp, D. T., and Souter, M. (1988). "A new rapid component in the cochlear response to brief electrical efferent stimulation," *Hear. Res.* **34**,

- Kemp, D. T., Ryan, S., and Bray, P. (1990). "A guide to the effective use of otoacoustic emissions," *Ear Hear.* **11**, 93–105.
- Kim, D. O. (1980). "Cochlear mechanics: Implications of electrophysiological and acoustical observations," *Hear. Res.* **2**, 297–317.
- Kim, D. O., Molnar, C. E., and Matthews, J. W. (1980). "Cochlear mechanics: Nonlinear behavior in two-tone responses as reflected in cochlear-nerve-fiber responses and in ear-canal sound pressure," *J. Acoust. Soc. Am.* **67**, 1704–1721.
- Kim, D. O., Dorn, P. A., Neely, S. T., and Gorga, M. P. (2001). "Adaptation of distortion product otoacoustic emission in humans," *J. Assoc. Res. Otolaryngol.* **2**, 31–40.
- Lieberman, 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.
- Lieberman, M. C. (1990). "Effect of chronic cochlear de-efferentation of auditory-nerve response," *Hear. Res.* **49**, 209–224.
- Long, G., Talmadge, C., Prieve, B., and Lahtinen, L. (2008). "Extraction of DPOAE generator and reflection components in the time domain in adults and infants," *Assoc. Res. Otolaryngol. Meeting*, Abstr. No. 180.
- Lonsbury-Martin, B. L., and Martin, G. K. (1990). "The clinical utility of distortion-product otoacoustic emissions," *Ear Hear.* **11**, 144–154.
- Matthews, J. W., and Molnar, C. E. (1986). "Modeling intracochlear and ear canal distortion products $2f_1 - f_2$," in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, New York), pp. 258–265.
- Neely, S. T., and Kim, D. O. (1983). "An active cochlear model showing sharp tuning and high sensitivity," *Hear. Res.* **9**, 123–130.
- Neely, S. T., and Kim, D. O. (1986). "A model for active elements in cochlear biomechanics," *J. Acoust. Soc. Am.* **79**, 1472–1480.
- Neely, S. T., and Kim, D. O. (2008). "Cochlear models incorporating active processes," in *Active Processes and Otoacoustic Emissions in Hearing*, edited by G. A. Manley, R. R. Fay, and A. N. Popper (Springer, New York), Chap. 11, pp. 381–394.
- Neely, S. T., Kim, D. O., and Gorga, M. P. (2003). "A novel method for making fast measurements of DPOAE as a continuous function of primary level," *Assoc. Res. Otolaryngol. Meeting*, Abstr. No. 361.
- Oxenham, A. J., and Shera, C. A. (2003). "Estimates of human cochlear tuning at low levels using forward and simultaneous masking," *J. Assoc. Res. Otolaryngol.* **4**, 541–554.
- Pickles, J. O. (1984). "Frequency threshold curves and simultaneous masking functions in single fibers of the guinea pig auditory nerve," *Hear. Res.* **14**, 245–256.
- Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (1991). "A review of otoacoustic emissions," *J. Acoust. Soc. Am.* **89**, 2027–2067.
- Puria, S. (2003). "Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions," *J. Acoust. Soc. Am.* **113**, 2773–2789.
- Ren, T., He, W., Scott, M., and Nuttal, A. L. (2006). "Group delay of acoustic emissions in the ear," *J. Neurophysiol.* **96**, 2785–2791.
- Rhode, W. S. (1971). "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique," *J. Acoust. Soc. Am.* **49**, 1218–1231.
- Rhode, W. S. (2007). "Basilar membrane mechanics in the 6–9 kHz region of sensitive chinchilla cochleae," *J. Acoust. Soc. Am.* **121**, 2792–2804.
- Robles, L., and Ruggero, M. A. (2001). "Mechanics of the mammalian cochlea," *Physiol. Rev.* **81**, 1305–1352.
- Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (1997). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," *J. Acoust. Soc. Am.* **101**, 2151–2163.
- Schairer, K. S., Ellison, J. C., Fitzpatrick, D., and Keefe, D. H. (2006). "Use of stimulus-frequency otoacoustic emission latency and level to investigate cochlear mechanics in human ears," *J. Acoust. Soc. Am.* **120**, 901–914.
- Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (1982). "Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique," *J. Acoust. Soc. Am.* **72**, 131–141.
- Shera, C. A. (2004). "Mechanisms of mammalian otoacoustic emission and their implications for the clinical utility of otoacoustic emissions," *Ear Hear.* **25**, 86–97.
- 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., and Zweig, G. (1991). "Reflection of retrograde waves within the cochlea and at the stapes," *J. Acoust. Soc. Am.* **89**, 1290–1305.
- Shera, C. A., Guinan, J. J., and Oxenham, A. J. (2002). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3318–3323.
- Shera, C. A., Tubis, A., and Talmadge, C. L. (2006). "Delays of SFOAEs and cochlear vibrations support the theory of coherent reflection filtering," *Assoc. Res. Otolaryngol. Meeting Abstr.*, Vol. **29**, revised poster No. 52.
- Siegel, J. H., Cerka, A. J., Recio-Spinoso, A., Temchin, A. N., van Dijk, P., and Ruggero, M. A. (2005). "Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering," *J. Acoust. Soc. Am.* **118**, 2434–2443.
- Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (1998). "Modeling otoacoustic emission and hearing threshold fine structure," *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.
- 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.
- Wilson, J. P., and Johnstone, J. R. (1975). "Basilar membrane and middle-ear vibration in guinea pig measured by capacitive probe," *J. Acoust. Soc. Am.* **57**, 705–723.
- Zurek, P. M., Clark, W. W., and Kim, D. O. (1982). "The behavior of acoustic distortion products in the ear canals of chinchillas with normal or damaged ears," *J. Acoust. Soc. Am.* **72**, 774–780.
- Zweig, G., Lipes, R., and Pierce, J. R. (1976). "The cochlear compromise," *J. Acoust. Soc. Am.* **59**, 975–982.
- 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. (1986). "Otoacoustic emissions' in a nonlinear cochlear hardware model with feedback," *J. Acoust. Soc. Am.* **80**, 154–162.
- Neely, S. T., and Stevenson, R. (2002). "SYSRES," Tech. Memo. 19, Boys Town National Research Hospital, Omaha, NE. (personal communication)