

# Obtaining reliable phase-gradient delays from otoacoustic emission data

Christopher A. Shera<sup>a)</sup>

*Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye & Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114*

Christopher Bergevin

*Department of Otolaryngology, Head and Neck Surgery, Columbia University, New York, New York 10032*

(Received 13 March 2012; revised 1 June 2012; accepted 6 June 2012)

Reflection-source otoacoustic emission phase-gradient delays are widely used to obtain noninvasive estimates of cochlear function and properties, such as the sharpness of mechanical tuning and its variation along the length of the cochlear partition. Although different data-processing strategies are known to yield different delay estimates and trends, their relative reliability has not been established. This paper uses *in silico* experiments to evaluate six methods for extracting delay trends from reflection-source otoacoustic emissions (OAEs). The six methods include both previously published procedures (e.g., phase smoothing, energy-weighting, data exclusion based on signal-to-noise ratio) and novel strategies (e.g., peak-picking, all-pass factorization). Although some of the methods perform well (e.g., peak-picking), others introduce substantial bias (e.g., phase smoothing) and are not recommended. In addition, since standing waves caused by multiple internal reflection can complicate the interpretation and compromise the application of OAE delays, this paper develops and evaluates two promising signal-processing strategies, the first based on time-frequency filtering using the continuous wavelet transform and the second on cepstral analysis, for separating the direct emission from its subsequent reflections. Altogether, the results help to resolve previous disagreements about the frequency dependence of human OAE delays and the sharpness of cochlear tuning while providing useful analysis methods for future studies.

© 2012 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4730916>]

PACS number(s): 43.64.Jb, 43.64.Bt, 43.64.Kc, 43.60.Hj [BLM]

Pages: 927–943

## I. INTRODUCTION

Some of the energy evoked from active processes in the inner ear leaks out into the ear canal and appears as sound (Kemp, 1978). These sounds, known as otoacoustic emissions (OAEs), provide a noninvasive window on the mechanics of the cochlea. Although emission levels serve as convenient assays of cochlear function, important information is also carried by emission phase (e.g., Shera and Guinan, 1999). For example, measurements of reflection-source OAE phase—or its time-domain counterpart, latency—have been used to probe mechanisms of emission generation (e.g., Zweig and Shera, 1995; Siegel *et al.*, 2005; Shera *et al.*, 2008; Bergevin and Shera, 2010; Meenderink and van der Heijden, 2010), to estimate the wavelength and delay of cochlear traveling waves (e.g., Neely *et al.*, 1988; Shera and Guinan, 2003; Moleti and Sisto, 2008; Harte *et al.*, 2009), to explore the effects of olivocochlear efferent feedback to the cochlea (e.g., Francis and Guinan, 2010), and to determine the characteristics of peripheral frequency selectivity in humans and other species (e.g., Shera *et al.*, 2002, 2010; Schairer *et al.*, 2006; Sisto and Moleti, 2007; Moleti *et al.*, 2008; Bergevin *et al.*, 2008; Lineton and Wildgoose, 2009; Bergevin *et al.*, 2010; Bentsen *et al.*, 2011; Bergevin, 2011; Joris *et al.*, 2011).

In many of these applications, the frequency gradient of reflection-source emission phase—the phase-gradient delay—is used to infer characteristics of mechanical responses within the cochlea, a possibility suggested by models of emission generation (Zweig and Shera, 1995; Bergevin and Shera, 2010). However, the interpretation of emission phase-gradient delays can be seriously confounded by their large and irregular fluctuations across frequency (i.e., by the delay counterpart of OAE macrostructure). At the signal-to-noise ratios (SNRs) typical of most OAE recordings from healthy ears, emission phase measurements are quite reproducible. Even though fluctuations due to noise are magnified by taking the derivative to obtain the phase-gradient delay, the great bulk of the observed variation is not generally due to measurement uncertainty. Rather, large fluctuations in OAE phase-gradient delays and their correlations with fluctuations in emission amplitude (Zweig and Shera, 1995; Talmadge *et al.*, 2000; Sisto *et al.*, 2007) are intrinsic to the emission process itself. Thus, the most troublesome and interesting issue—the one that cannot be addressed simply by improving the measurement frequency resolution or by increasing the number of averages—is not the reliability of the numerical differentiation needed to compute the delay but the fact that the emission itself is intrinsically irregular. The coherent-reflection model indicates that the frequency fluctuations in emission magnitude and phase trace their origin to spatial irregularities in cochlear micromechanics that give rise to reverse-traveling waves. Evidently, micromechanical

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: shera@epl.meei.harvard.edu

irregularities are both a blessing and a curse: Although essential to the process of emission generation that permits noninvasive exploration of cochlear mechanics, irregularities also introduce fluctuations in the frequency response that obscure the information one hopes to recover.<sup>1</sup>

A variety of strategies have been adopted for extracting phase-gradient delays while mitigating the impact of intrinsic fluctuations and measurement noise. These include fitting trend lines to delay data pooled across subjects after excluding measurements with low SNR (e.g., [Shera and Guinan, 2003](#)), computing SNR-weighted averages across frequency (e.g., [Lineton and Wildgoose, 2009](#)), and smoothing the measured phase by fitting regression lines (e.g., [Francis and Guinan, 2010](#)) or smoothing splines (e.g., [Schairer et al., 2006](#)) before computing the gradient. Although each of these strategies seems *a priori* reasonable, they do not all yield the same result. For example, smoothing the phase produces delay trends that differ systematically from those obtained without smoothing ([Schairer et al., 2006](#); [Bentsen et al., 2011](#)). Although [Schairer et al. \(2006\)](#) emphasize the benefits of the smoothing procedure—it allows inclusion of additional data with lower SNR—the optimal method for obtaining delay trends has not been established.

Here we address the problem of determining reliable and physically meaningful phase-gradient delay trends from reflection-source OAE data. Our approach is to analyze simulated but realistic data obtained from a phenomenological model of the emission process. Because the parameters of the model are known, we can evaluate different procedures and determine which yield the most reliable estimates of the underlying model trend.<sup>2</sup> In addition to exploring existing methods (namely, phase smoothing, energy-weighting, and SNR-based data exclusion), we describe and evaluate two novel ones (all-pass factorization and peak-picking). The first of these, motivated by hints in the literature that macrostructure of stimulus-frequency OAEs (SFOAEs) might be minimum-phase, proves unreliable. The second, however, not only performs well but also helps explain why other strategies succeed or fail. We also show that estimates of emission latencies can be distorted by multiple internal reflection and the buildup of standing waves inside the cochlea. Obtaining reliable estimates of the delays associated with cochlear filtering requires that these effects be removed from the data, and we develop two new strategies—one based on time-frequency filtering and the other on cepstral analysis—for separating the direct emission from its subsequent reflections.

## II. GENERAL METHODS

### A. Simulating emissions

We represent the measured SFOAE pressure at frequency  $f$  as an OAE signal with additive noise as follows:

$$P_{\text{SFOAE}}(f) = \hat{P}_{\text{SFOAE}}(f) + N(f). \quad (1)$$

When multiple internal reflection within the cochlea can be neglected (see Sec. V), the model emission pressure has the following form ([Shera, 2003](#)):

$$\hat{P}_{\text{SFOAE}} = P_0 G_{\text{ME}} R, \quad (2)$$

where  $P_0(f)$  is the stimulus source pressure,  $G_{\text{ME}}(f)$  characterizes roundtrip middle-ear transmission, and  $R(f)$  is the cochlear reflectance, representing the complex amplitude of the reverse-traveling wave (normalized by the ingoing wave) at the stapes. We describe the production of reverse-traveling waves within the cochlea using an equation borrowed from the coherent-reflection model of reflection-source OAE generation ([Shera et al., 2005, 2008](#)):

$$R(f) \sim \int \epsilon(x) \mathcal{W}^2(x, f) dx, \quad (3)$$

where  $\epsilon(x)$  represents the micromechanical irregularity and  $\mathcal{W}(x, f)$  is a weighting function summarizing fluid-membrane coupling and roundtrip pressure-difference wave propagation between the stapes and the site of scattering at cochlear position  $x$ . Although the integral sums contributions to the emission from wavelets scattered throughout the cochlea, the emission at any given frequency is dominated by wavelets originating from the peak region of the traveling wave. Near the peak, located at  $\hat{x}(f)$ , we approximate  $\mathcal{W}(x, f)$  by a Gaussian envelope and a locally linear phase, a phenomenological description previously used to capture the essential features of the coherent-reflection model ([Zweig and Shera, 1995](#); [Talmadge et al., 2000](#)),

$$\mathcal{W}(x, f) = \hat{\mathcal{W}} e^{-[(x-\hat{x})/2\Delta x]^2} e^{-2\pi i(x-\hat{x})/\hat{\lambda}}. \quad (4)$$

The parameters  $\hat{\lambda}$  and  $\Delta x$  determine, respectively, the local wavelength and spatial spread of the traveling-wave envelope. To approximate the variation in tuning, wavelength, and delay believed characteristic of the human cochlea (e.g., [Shera and Guinan, 2003](#)), we allow the parameters to vary with location [or, equivalently with the local characteristic frequency, which we take to vary exponentially with position,  $\text{CF}(x) = \text{CF}(0)e^{-x/l}$ , with  $\hat{x}(f)$  satisfying  $\text{CF}(\hat{x}(f)) = f$ , and where  $l$  is the space constant of the tonotopic map; in humans,  $\text{CF}(0) \cong 20$  kHz and  $l \cong 7.2$  mm ([Greenwood, 1990](#))]. Specifically, for the wavelength we take  $\hat{\lambda}(\text{CF}) = l/N_W(\text{CF})$ , with  $N_W(\text{CF}) = 5.5 \cdot (\text{CF}/\text{kHz})^\gamma$  with  $\gamma = 0.5$  for  $\text{CF} < 1$  kHz and  $\gamma = 0.37$  for  $\text{CF} > 1$  kHz. In practice, we eliminate the discontinuity in  $\gamma$  by changing its value smoothly over a span of a few hundred hertz. For the spatial spread of the wave, we take  $\Delta x = l/\sqrt{2\pi Q_{\text{ERB}}}$ , with  $Q_{\text{ERB}}(\text{CF}) = 10 \cdot (\text{CF}/\text{kHz})^{0.3}$ .

When the irregularity function  $\epsilon(x)$  contains spatial frequencies near  $2/\hat{\lambda}$  (e.g., when the irregularities are truly irregular or random), coherent-reflection theory indicates that the resulting emissions manifest a mean phase-gradient delay<sup>3</sup> given by  $\bar{\tau}_{\text{SFOAE}}(f) \cong 2l/\hat{\lambda}f$  ([Zweig and Shera, 1995](#)). For the parameters used here, the expected emission delay expressed in periods of the emission frequency ( $\bar{N}_{\text{SFOAE}} \equiv f\bar{\tau}_{\text{SFOAE}}$ ) therefore becomes<sup>4</sup>

$$\bar{N}_{\text{SFOAE}}(f) = 2N_W(\text{CF})|_{\text{CF}=f} = 11 \cdot (f/\text{kHz})^\gamma. \quad (5)$$

We opt to express the emission delay in dimensionless form to facilitate comparison with other dimensionless parameters

of hearing, such as the quality factors that measure the sharpness of cochlear tuning. Because the power-law exponents characterizing the variation of  $Q_{\text{ERB}}(\text{CF})$  and  $\bar{N}_{\text{SFOAE}}(f)$  are nonzero, the model incorporates deviations from scaling similar to those found in otoacoustic and neural measurements (e.g., Shera and Guinan, 2003). The change in the power-law exponent  $\gamma$  simulates the “bend” observed in human  $N_{\text{SFOAE}}(f)$  curves near 1–1.5 kHz (Shera and Guinan, 2003).<sup>5</sup> The frequency where this bend occurs provides an otoacoustic estimate of the location of the transition between apical-like and basal-like behavior in the cochlea (Shera *et al.*, 2010).

Computations of the reflectance using the integral in Eq. (3) were performed numerically by partitioning the cochlea into 3500 longitudinal segments. SFOAEs for different *in silico* subjects were obtained by varying  $\epsilon(x)$  from subject to subject using different samples of Gaussian spatial noise; all other model parameters were fixed across subjects. We fixed the root-mean-square (rms) value of the irregularity function  $\epsilon(x)$  at 0.03 and then exploited the linearity of the model by taking  $G_{\text{ME}} = \hat{\mathcal{V}} = 1$  and adjusting the stimulus amplitude  $P_0$  to yield a mean emission level of  $\sim 0$  dB SPL. After computing  $\hat{P}_{\text{SFOAE}}(f)$  using Eq. (2), we added a complex-valued, Gaussian noise component  $N(f)$  at each frequency to obtain  $P_{\text{SFOAE}}(f)$  from Eq. (1). The rms noise level was adjusted so that the ratio  $|\hat{P}_{\text{SFOAE}}/N|^2$ , averaged across frequency, equaled the desired mean SNR. Unless otherwise specified, we employed a mean SNR of 15 dB.

We computed SFOAEs at frequencies spanning 0.4–8 kHz (4.3 octaves) with a frequency resolution of 65 points/octave (Schaerer *et al.*, 2006). Except when the emission fell into the noise, or when an SFOAE magnitude notch and associated phase transition was exceptionally sharp, this frequency spacing proved sufficient to resolve ambiguities in phase unwrapping, and simulations performed with higher resolution gave similar results. Phase-gradient delays were computed from the unwrapped phase using three-point centered differences (e.g., Press *et al.*, 2007, Sec. 5–7). In some cases (described in the following), the unwrapped phase was smoothed prior to taking the derivative.

## B. Loess smoothing

Trend lines to  $N_{\text{SFOAE}}$  data were computed using robust loess smoothing (Cleveland, 1993, Sec. 3.2). Loess smoothing is a nonparametric, statistical procedure that employs local regression to find a smooth curve that captures the secular variation or trend of “noisy” or irregular data. Because we plot  $N_{\text{SFOAE}}$  versus frequency data on logarithmic axes, both coordinates were log-transformed before finding the trend. Log-transforming the ordinate is necessary to equalize the variance of the data across frequency. Although computing  $\log N_{\text{SFOAE}}$  necessitates eliminating non-positive values, the discarded data were invariably outliers with negligible influence on the trend (as assayed using fits performed without the log transformation). The loess smoothing parameters  $\lambda_{\text{loess}}$  and  $\alpha_{\text{loess}}$  specify, respectively, the degree of the local fitting polynomial and the size of the moving window as a fraction of the total number of data points. We used  $\lambda_{\text{loess}} = 1$

and  $\alpha_{\text{loess}} = 0.2$ , which at any given frequency produces a locally linear regression based on the nearest 20% of the data. Because we used data taken at log-spaced frequency intervals and performed loess smoothing on log-transformed coordinates, a constant value of the smoothing parameter  $\alpha_{\text{loess}}$  produces a moving window that spans a constant number of octaves (in our case,  $0.2 \times 4.3 = 0.86$  octaves).<sup>6</sup> Unless otherwise specified, the loess fits were performed using uniform weights (in some cases, they were energy-weighted). We circumvented potential singularities in the fitting procedure caused by multiple data points at the same abscissa by randomly dithering the frequencies by 0.1%. Confidence intervals on the trend were computed using bootstrap resampling.

## III. SIMULATED EMISSIONS

Figure 1 shows example computations of  $P_{\text{SFOAE}}(f)$  obtained for two different *in silico* subjects using Eq. (1) and standard parameter values. The simulated emissions reproduce the major features of measured SFOAEs, including their notchy magnitude spectra and rapidly rotating phases (e.g., Shera and Guinan, 1999; Siegel *et al.*, 2005). Wobbles in the phase correlate with undulations and notches in

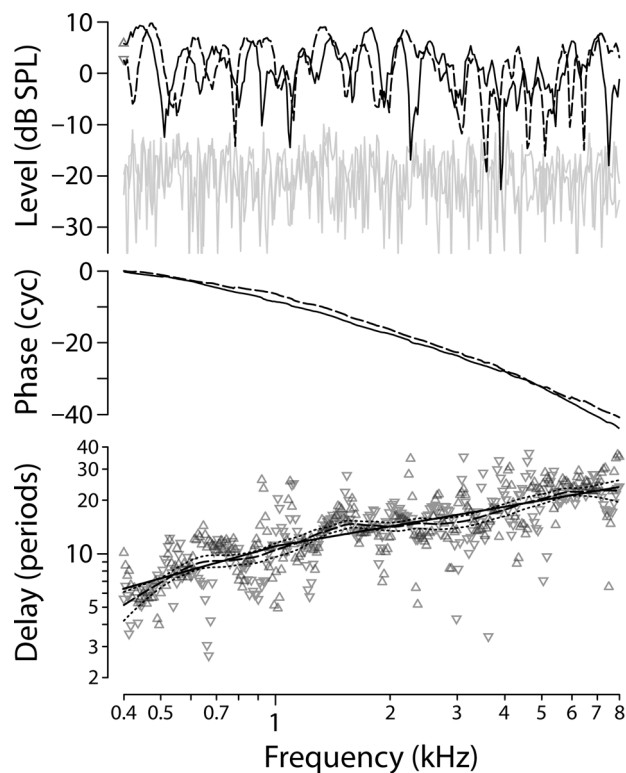


FIG. 1. Simulated SFOAEs in two *in silico* subjects. Top to bottom, the three panels show model SFOAE level, unwrapped phase, and delay in periods ( $N_{\text{SFOAE}}$ ) computed using Eq. (1) for two different irregularity patterns (i.e., subjects) using standard parameters. The two subjects are shown using different line types (solid and dashed). Delays shown using the regular ( $\Delta$ ) and inverted triangles ( $\nabla$ ) were computed from the solid and dashed phase curves, respectively. The model trend used in the simulations,  $\bar{N}_{\text{SFOAE}}(f)$ , is shown in the bottom panel (solid line). The bottom panel also gives the loess trend (dashed line) and its 95% confidence intervals (dotted lines) computed from the pair of subjects. In this example, the mean SNR is 21 dB.

emission magnitude, as expected from constraints imposed by causality (Sisto *et al.*, 2007). Although barely visible in the phase itself on this scale—note that the phase falls through 40 cycles over the frequency range of the figure—the impact of phase wobbles is readily apparent in the phase-gradient delay. Although the model values of  $N_{\text{SFOAE}}(f)$  generally increase with frequency, as expected from the trend  $\bar{N}_{\text{SFOAE}}(f)$  predicted by coherent-reflection theory, there are major local deviations, sometimes as large as a few hundred percent. Some of this variability arises from the additive measurement noise,  $N(f)$ , whose effects are magnified by the frequency derivative taken to compute the delay. But at all but the lowest SNRs, most of the variance about the overall trend reflects the randomness inherent in the process of emission generation itself (i.e., through the irregular distribution of micromechanical perturbations responsible for scattering the wave).

Figure 1 thus illustrates the problem addressed in this paper: How best to analyze measurements of  $P_{\text{SFOAE}}(f)$  to reduce the effects of noise and the inherent variability in emission magnitude and phase in order to reliably estimate characteristics of the mechanical response (e.g., the wavelength and delay of the traveling wave).

#### IV. EVALUATING THE ANALYSIS METHODS

We used computer simulations to evaluate six different methods for estimating the underlying model delay trend,  $\bar{N}_{\text{SFOAE}}(f)$  from SFOAE measurements. These six methods included both previously published procedures (e.g., phase smoothing, energy-weighting, SNR-based data exclusion), novel strategies (e.g., peak-picking, all-pass factorization), and a control (nothing special). The rationale and implementation details for all but one of the methods are outlined in separate upcoming sections, following a brief presentation of the results. The method left undescribed is the control method identified as “nothing special.” Living up to its name, this method is easy to summarize: no special signal processing, data selection, or massaging was performed, and the loess trends were computed using uniform weighting and all the available data, regardless of SNR.

To perform the evaluations, we generated SFOAEs for 1500 different *in silico* subjects (i.e., different irregularity patterns), randomly divided them into 100 groups of 15, and computed phase-gradient delays for every subject using each method. For each method, we pooled the delays across subjects within each group and computed loess trend lines to find the group trend,  $N_{\text{SFOAE}}(f)$ . The percent estimation error (perr) for each method in each group was obtained using the formula  $\text{perr}(f) = 100[1 - N_{\text{SFOAE}}(f)/\bar{N}_{\text{SFOAE}}(f)]$ . Finally, the mean and standard deviations of  $\text{perr}(f)$  were computed across the 100 groups.

Figures 2–5 summarize our analysis of the six methods for obtaining group  $N_{\text{SFOAE}}(f)$  trends. As shown in Figs. 2 and 3, the majority of the methods produced reliable estimates of the underlying model trend,  $\bar{N}_{\text{SFOAE}}(f)$ , with estimation errors and their standard deviations averaging less than a few percent. In all cases, the variance of the estimation error increases at the extremes of the measured

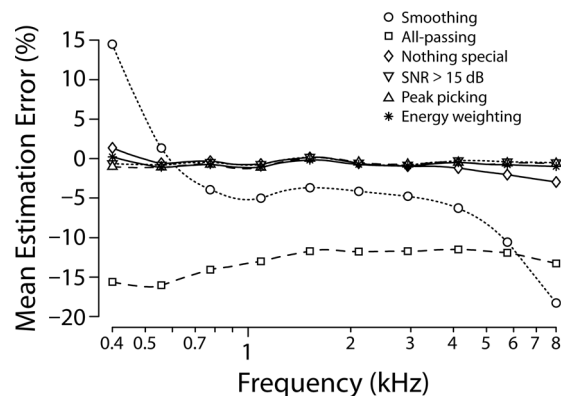


FIG. 2. Mean percent estimation errors versus frequency for six methods of computing group  $N_{\text{SFOAE}}(f)$  trends. The mean error shown here was obtained by averaging the percent estimation error,  $\text{perr}(f)$ , across 100 groups. In each group, the trend was computed using pooled simulated data from 15 *in silico* subjects. The same subjects and groups were used to evaluate each method. The mean SNR in the simulations was set to 15 dB.

frequency range (Fig. 3), where the loess smoothing procedure is least (and asymmetrically) constrained. Two of the six methods (phase smoothing and all-pass factorization), however, produced significant systematic error. The phase-smoothing method, for example, manifests a frequency-dependent bias, overestimating the delay at low frequencies and underestimating it at medium and high.

Figure 4 shows how the estimation error varies with the mean SNR employed in the simulations. With the exception, again, of the phase-smoothing and all-pass factorization methods, the trends across SNR appear similar. At high SNRs, the SNR-based selection criterion has no effect, and the method becomes equivalent to doing “nothing special.” At mean SNRs lower than the 15 dB criterion, the selection becomes too restrictive and the method begins to fail; eventually the method cannot be evaluated without lowering the criterion. (If the selection criterion were lowered enough, the method would again become equivalent to doing nothing special.) Overall, the peak-picking and energy-weighting

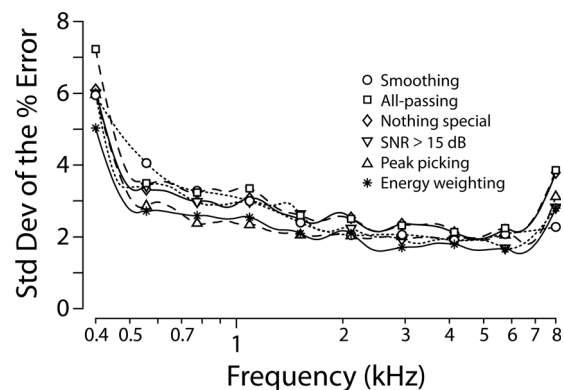


FIG. 3. Standard deviations of the percent estimation errors shown in Fig. 2. Results are shown versus frequency for the six methods of computing  $N_{\text{SFOAE}}(f)$  trends. At each frequency, standard deviations were computed across the 100 groups of 15 subjects. The variability of the estimates is largest near the edges of the data (i.e., at the low- and high-frequency extremes), where the loess smoothing procedure is least constrained. The approximate span of the 95% confidence intervals for the trends shown in Fig. 2 can be computed by multiplying the standard deviations shown here by 0.4 ( $= 2 \times 2/\sqrt{100}$ ).

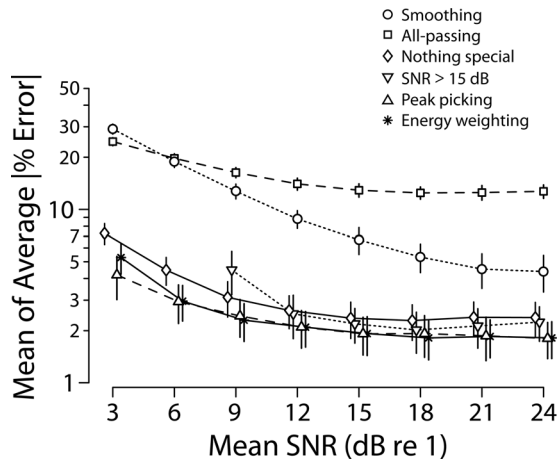


FIG. 4. Means and standard deviations of the frequency-averaged estimation error versus the mean SNR employed in the simulation. At each SNR, percent estimation errors for each group, computed as indicated in Fig. 2, were averaged across frequency and the mean of the result then computed across the 100 groups. Results are shown for each of the six methods of computing group  $N_{\text{SFOAE}}(f)$  trends; as before, the same subjects and groups were used to evaluate each method. All SNRs are multiples of 3 dB; data points whose symbols or error bars would otherwise overlap have been slightly offset from one another along the abscissa for clarity. Note that for SNR < 9 dB, too many data points were discarded to evaluate the method involving the SNR selection criterion of 15 dB.

methods perform best, with peak-picking winning out at the lowest SNR.

To explore the dependence on the number of subjects in the group, we varied the number of subjects ( $N_{\text{subj}}$ ) used to compute the group trends and then found the mean absolute deviation of the resulting estimation error, averaged across groups and frequency. The statistic assesses the variability of the estimation error and thus measures the uncertainty in the group trend, whatever that trend may be. Results are shown in Fig. 5. For all but the phase-smoothing method, whose mean precision plateaus, the uncertainty of the group

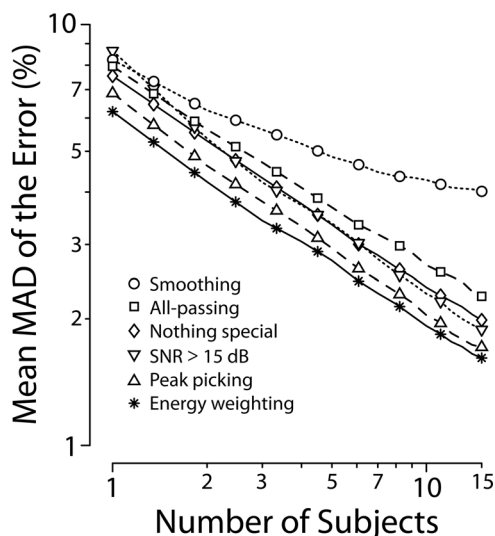


FIG. 5. Mean absolute deviations (MADs) of the percent estimation errors for the six ways of computing group  $N_{\text{SFOAE}}(f)$  trends. The MADs, averaged across frequency, are shown as a function of the number of subjects ( $n$ ) used to compute each of the 100 group trends. The mean SNR was set to 15 dB.

trend decreases as  $N_{\text{subj}}^{-1/2}$  as the number of subjects in the group increases.

Combining the results from Figs. 2–5, the top overall performers are peak-picking and energy-weighting. Details of the various methods are presented in the following sections.

### A. Peak-picking

We begin by describing an algorithm we dub “peak-picking,” which performs well and helps explain the success of other methods. The method was motivated by our observation that the model values of  $N_{\text{SFOAE}}(f)$  lying closest to the expected trend often occur at frequencies near local maxima in emission level. Figure 6 illustrates this correlation in an individual subject. The gray dots in Fig. 6(B) show values of the delay  $N_{\text{SFOAE}}(f)$  obtained from the model computations of  $P_{\text{SFOAE}}(f)$  level and phase reproduced in Fig. 6(A); the solid line in Fig. 6(B) gives the expected trend,  $\bar{N}_{\text{SFOAE}}(f)$ . Data points that fall within 10% of the trend are highlighted in black. The black dots in Fig. 6(A) mark data at the same frequencies as those identified in Fig. 6(B) based on proximity to the delay trend,  $\bar{N}_{\text{SFOAE}}(f)$ . In the plot of  $|P_{\text{SFOAE}}|$  shown in the top of Fig. 6(A), most (although not all) of the

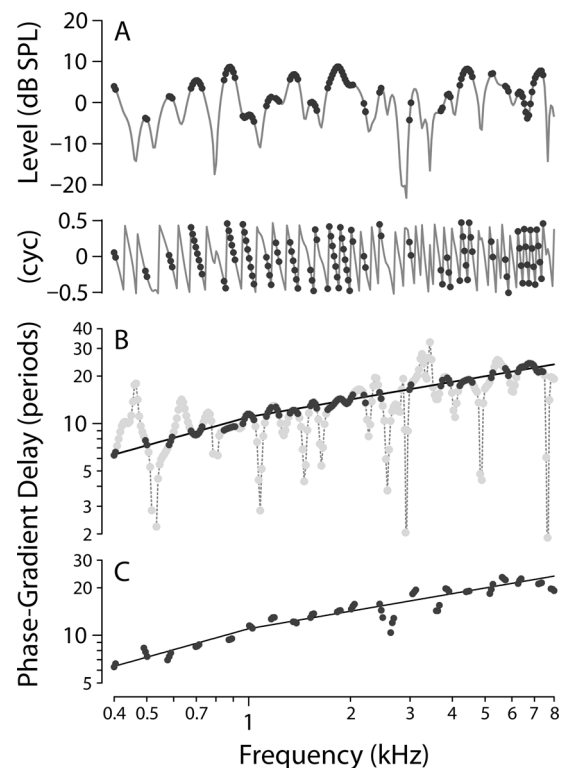


FIG. 6. Basis for and results of the peak-picking algorithm. (A) Example simulated SFOAE level curve (top) and wrapped phase (bottom). Black dots mark data points determined using the delay values in (B). (B) Emission phase-gradient delay in periods,  $N_{\text{SFOAE}}$  (gray and black dots), computed from the unwrapped phase. The solid line gives the trend,  $\bar{N}_{\text{SFOAE}}$ , used in the model computations. The black dots, here and in (A), identify data at frequencies where the value of  $N_{\text{SFOAE}}$  lies within 10% of the model trend,  $\bar{N}_{\text{SFOAE}}$ . (C) Values of  $N_{\text{SFOAE}}$  (black dots) selected by the peak-picking algorithm. The data points are those from panel B that occur at frequencies straddling local maxima in SFOAE level. Each peak is represented by three points—the maximum itself and one value on either side. The model trend is shown for comparison (solid line).

black dots lie close to peaks in emission magnitude; in the plot of wrapped phase shown in the bottom of Fig. 6(A), the black dots generally line up along the more locally linear segments of the curve, skipping over regions of significant phase curvature (e.g., those correlated with notches in the magnitude spectrum).

The peak-picking method for extracting phase-gradient delays inverts the procedure used to select the black dots in Figs. 6(A) and 6(B). Rather than using proximity to the delay trend to mark emission peaks, the method uses peak locations to identify delays  $N_{\text{SFOAE}}$  likely to be near the trend. Thus, the algorithm considers only those values of  $N_{\text{SFOAE}}(f)$  that occur near frequencies corresponding to peaks in  $|P_{\text{SFOAE}}(f)|$ ; other data points are ignored when computing  $N_{\text{SFOAE}}(f)$  trends. For the example illustrated in Fig. 6, the  $N_{\text{SFOAE}}$  data selected using this method are shown in Fig. 6(C). In the implementation tested here, the selection included the three data points straddling a peak frequency (i.e., the peak itself and the points on either side). To reduce the impact of noise on the identification of peak frequencies, emission levels were gently smoothed using Savitzky–Golay filters (e.g., Press *et al.*, 2007, Sec. 14–9) prior to locating the maxima. Smoothing generally had only modest effects on the results—peaks, by virtue of being peaks, tend to have better SNRs than other regions of the emission spectrum.

## B. SNR-based exclusion criteria and energy-weighting

The success of the peak-picking strategy suggests that other procedures whose net effect is to emphasize delays near magnitude maxima may also perform well. Among these are popular strategies such as applying data-exclusion criteria based on local SNR [e.g., ignoring delay values at frequencies where SFOAE level is less than a specified amount above the noise floor (e.g., Shera and Guinan, 2003; Bergevin, 2011)] and computing energy-weighted delays [in which delay values are weighted in proportion to a measure of the local emission “energy” and then averaged over some frequency band or group of subjects (e.g., Lineton and Wildgoose, 2009; Bentsen *et al.*, 2011)]. For our tests of the SNR-based exclusion method, we set the criterion so that delays at frequencies with  $\text{SNR} < 15$  dB—the mean SNR employed in the computation of  $P_{\text{SFOAE}}$ —were excluded when computing the loess trend.<sup>7</sup> The exclusion criterion was thus applied as late in the process as possible (i.e., to the delay values rather than to the phase before the derivative was computed). In our version of energy-weighting, the weighting by local SFOAE energy was performed during the loess smoothing. Delay values were weighted by the value  $|P_{\text{SFOAE}}/P_{\text{REF}}|^2$ , where  $P_{\text{REF}}$  was taken as the mean value of  $|P_{\text{SFOAE}}|$  over the one-octave range about each data point.

## C. Phase smoothing

Noise and other fluctuations in the phase-gradient delay can be substantially reduced by smoothing the phase prior to taking the derivative. We followed the procedure outlined by Schairer *et al.* (2006) and smoothed the phase using smoothing splines as implemented in MATLAB’s csaps function (The MathWorks, Natick, MA). As in Schairer *et al.* (2006), data

with SNR less than 6 dB were excluded from the fitting procedure but were otherwise weighted by their SNR; all splines were computed with respect to  $\log f$ . The amount of smoothing is controlled by a parameter,  $s$ , whose value ( $0 \leq s \leq 1$ ) specifies the relative weight given to reducing the global (integrated) squared curvature of the fit and to minimizing local weighted deviations from the data points.<sup>8</sup> At the extremes, the value  $s = 0$  (no smoothing) gives a standard cubic spline that passes through every data point, and the value  $s = 1$  (maximum smoothing) performs a linear regression to obtain the best fit with zero curvature (i.e., a straight line). The value of  $s$  that one chooses depends on assumptions, prior knowledge, or conviction about how smooth the fit should be. Schairer *et al.* explored a range of possible values and settled on the value  $s = 0.9$  as appropriate for their data. With this value, the smoothing retains the curve’s large-scale features but irons out much of the microstructure and noise (cf. Schairer *et al.*, 2006, Fig. 5). In simulations performed at similar SNR, but extending over a slightly larger frequency range, we found that this same value of  $s$  yielded visually comparable amounts of smoothing for our data.

Unfortunately, smoothing the phase proved generally undesirable. Figure 7 shows how even modest phase smoothing (top panel) can bias estimates of phase-gradient delay (bottom panel). Although the smoothing procedure maintains the overall shape of the phase—indeed, the raw and smoothed phase curves are nearly indistinguishable when viewed on a scale encompassing the full frequency range (0.4–8 kHz) of the data—the smoothing not only eliminates excursions due to phase microstructure (e.g., the delay notch near 3.3 kHz), but also distorts the gradient near SFOAE magnitude peaks (dots), where the delay is generally most representative of the trend. The net result is to preserve the global shape of the phase curve while significantly underestimating the delay trend actually used in the model computations (dotted line). The extent of the bias depends, of course, on the amount of smoothing performed (i.e., on  $s$ ), as well as

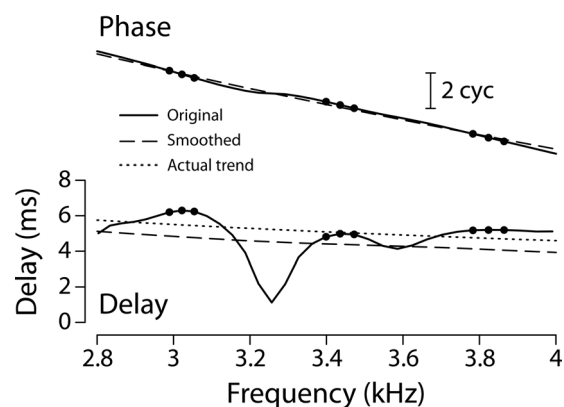


FIG. 7. Effects of phase smoothing on phase-gradient delay. (Top) A segment of simulated SFOAE phase both before (solid line) and after smoothing (dashed line). Triplets of dots indicate frequencies straddling local maxima in SFOAE magnitude (as might be used in the peak-picking algorithm). (Bottom) The corresponding values of phase-gradient delay. The dotted line shows the actual delay trend,  $\bar{\tau}_{\text{SFOAE}}(f)$ , used in the model computations. The phase smoothing was performed on the entire 0.4–8 kHz phase curve using a smoothing parameter of  $s = 0.9$ .

on factors that affect the integrated curvature of the phase (e.g., the total frequency span of the data and the amount of measurement noise). Previous applications of phase smoothing to real data found that the resulting delay estimates were significantly shorter than those obtained without smoothing (Schairer *et al.*, 2006; Bentsen *et al.*, 2011), consistent with these results.

#### D. Smoothing by all-pass factorization

This method is based on the possibility that the spectral notches and phase wobbles that constitute SFOAE macrostructure can be “divided out” to recover a smoothly varying delay. To be effective, the method requires that SFOAE macrostructure be minimum-phase, at least to a useful approximation. We know that stimulus-frequency OAEs are causal functions (Shera and Zweig, 1993). In the time domain, causality requires that the emission not precede the stimulus. In the frequency domain, causal functions can be written as the product of a minimum-phase and an all-pass component consisting of a (possibly frequency-dependent) delay (e.g., Papoulis, 1962, Sec. 10–3). Applying this factorization to SFOAEs, we let

$$P_{\text{SFOAE}}(f) \equiv P_{\text{MP}}(f)e^{i\phi_{\text{AP}}(f)}, \quad (6)$$

where  $P_{\text{MP}}(f)$  is minimum-phase and the all-pass component consists of a phasor with phase  $\phi_{\text{AP}} = -2\pi \int \tau_{\text{AP}}(f)df$  defined by the delay,  $\tau_{\text{AP}}(f)$ . Because  $P_{\text{MP}}(f)$  is minimum-phase, the logarithm of its magnitude ( $\log|P_{\text{MP}}| = \log|P_{\text{SFOAE}}|$ ) and its phase ( $\phi_{\text{MP}} \equiv \angle P_{\text{MP}}$ ) are not independent; one can be computed from the other using the Hilbert transform. As a consequence, spectral variations in  $\log|P_{\text{MP}}(f)|$  (i.e., the macrostructure in SFOAE level) have necessary counterparts in the phase  $\phi_{\text{MP}}$  and its phase-gradient delay ( $2\pi\tau_{\text{MP}} = -d\phi_{\text{MP}}/df$ ).

Note that if the troublesome phase transitions and wobbles evident in  $\phi_{\text{SFOAE}}$  were largely confined to  $\phi_{\text{MP}}$ , then one could remove their influence on  $\tau_{\text{SFOAE}}$  by simply subtracting them out prior to computing the phase-gradient delay (i.e., by computing the derivative of  $\phi_{\text{SFOAE}} - \phi_{\text{MP}}$  rather than of  $\phi_{\text{SFOAE}}$ ). In other words, one could potentially reduce or eliminate the variance due to spectral macrostructure by extracting the all-pass component of  $P_{\text{SFOAE}}$ , whose phase  $\phi_{\text{AP}} = \phi_{\text{SFOAE}} - \phi_{\text{MP}}$  and delay  $\tau_{\text{AP}} = \tau_{\text{SFOAE}} - \tau_{\text{MP}}$  would (ideally, and at sufficient SNR) be relatively smooth functions of frequency. As empirical support for this suggestion, several studies have noted correlated variations in SFOAE spectral level and group delay (Siegel *et al.*, 2005; Sisto *et al.*, 2007); correlated fluctuations in level and delay are also evident in the simulated SFOAEs shown in Fig. 1. Thus, at least some of the spectral fluctuations in  $\tau_{\text{SFOAE}}$  appear consistent with the minimum-phase behavior required by the proposed analysis method.

Although the all-pass factorization method appears *a priori* promising, the estimation errors reported in Fig. 2 indicate that the method fails in practice. Figure 8 provides an example showing that even at high SNR, the delay,  $\tau_{\text{AP}}$ , computed from the all-pass component of the SFOAE,<sup>9</sup> is generally no smoother than the delay  $\tau_{\text{SFOAE}}$  itself. Although

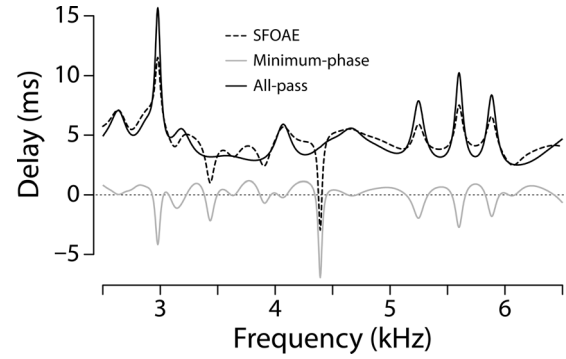


FIG. 8. SFOAE delay and its minimum-phase and all-pass components. The dashed line shows SFOAE phase-gradient delay  $\tau_{\text{SFOAE}}(f)$  computed from  $P_{\text{SFOAE}}(f)$  in an example *in silico* subject with no added noise (SNR > 200 dB). The gray line shows the delay  $\tau_{\text{MP}}$  of the minimum-phase component and the black line shows the delay  $\tau_{\text{AP}}$  of the all-pass component. Note that  $\tau_{\text{SFOAE}} = \tau_{\text{MP}} + \tau_{\text{AP}}$ . The minimum-phase and all-pass components were computed using Hilbert transforms, as implemented in MATLAB’s RCEPS function.

the method smooths out some phase wobbles (e.g., in the region near 3.5 or 4.5 kHz in Fig. 8), it accentuates others (e.g., near 5.5 kHz). Overall the method provides little smoothing and introduces a bias in the delay trend.

#### V. MITIGATING THE EFFECTS OF MULTIPLE INTERNAL REFLECTION

When they encounter the impedance mismatch at the cochlear boundary with the middle ear, reverse-traveling waves are partially reflected back into the cochlea, where they can serve as the stimulus for additional reemission (e.g., Shera and Zweig, 1991; Dhar *et al.*, 2002). This iterated process of multiple internal reflection gives rise to a succession of emission components with longer and longer delays. Under many circumstances, these higher-order contributions to the total emission can be neglected, and the discussion so far has assumed that the measured SFOAE is dominated by waves emitted directly from the region of generation. In this case, the emission pressure  $\hat{P}_{\text{SFOAE}}$  is well approximated by Eq. (2). However, when  $|R|$  is large—as it might be at low stimulus levels in strongly emitting subjects—multiple-internal reflections can make a significant contribution to the SFOAE. More generally,  $\hat{P}_{\text{SFOAE}}$  has the form (Shera, 2003)

$$\hat{P}_{\text{SFOAE}} = \frac{P_0 G_{\text{ME}} R}{1 - RR_{\text{stapes}}} \cong \hat{P}_{\text{SFOAE}}^0 \times \left( 1 + RR_{\text{stapes}} + (RR_{\text{stapes}})^2 + \dots \right), \quad (7)$$

where  $R_{\text{stapes}}(f)$  is the stapes reflection coefficient for retrograde cochlear waves and  $\hat{P}_{\text{SFOAE}}^0$  represents the component due to direct emission. The infinite series, which converges for  $|RR_{\text{stapes}}| < 1$ , encapsulates the process of multiple reflection.

By modifying the emission amplitude and phase, these additional terms, when significant, can bias efforts to determine emission delays and relate them to underlying processes of cochlear mechanics. The most serious problems occur at frequencies where the phase of  $RR_{\text{stapes}}$  is close to 0 (mod  $2\pi$ ). At such frequencies, higher-order reflections

combine in phase with one another, creating a peak in  $|\dot{P}_{\text{SFOAE}}|$  and, in the limiting case, a spontaneous emission or SOAE (Shera, 2003). Figure 9 illustrates magnitude peaks arising from multiple internal reflection using simulated values of  $P_{\text{SFOAE}}$  computed using parameters that yield a relatively large mean value of  $|RR_{\text{stapes}}|$  to highlight the effect. At these same peak frequencies, the phase-gradient delay becomes anomalously large, reflecting the fact that the re-emission process is spread out over a long time (multiple roundtrips). In the limiting case when the multiple reflections give rise to self-sustaining intracochlear standing waves (SOAEs), the “group delay” of the evoked emission becomes effectively infinite. If naively applied, normally effective analysis procedures, such as peak-picking or energy-weighting, will therefore overestimate the actual phase-gradient delays of interest, which are those of the direct emission,  $P_{\text{SFOAE}}^0$ , unbiased by multiple internal reflection. In this section we develop two methods for removing the confounding effects of multiple reflection from SFOAE measurements.

### A. Time-frequency prefiltering using the CWT

Wavelet transforms can be used both to visualize multiple reflections across time and frequency and to remove them from the measurements. The continuous wavelet transform (CWT) is a linear time-frequency analysis tool (Morlet et al., 1982) that decomposes signals into a superposition of “daughter” wavelets, each consisting of frequency-scaled and time-shifted copies of an underlying oscillatory, wave-like function known as the “mother” wavelet, whose form can be chosen to fit the application. Like other time-frequency analysis techniques, such as the short-time Fourier transform (STFT), the CWT provides information on the time evolution of the different frequency components of the response. Unlike the STFT, the duration of the analysis window is not fixed, but varies with frequency, providing improved time resolution at high frequencies and improved frequency resolution at low frequencies. Several helpful

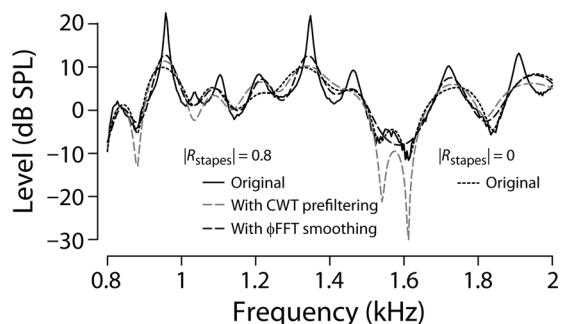


FIG. 9. Simulated SFOAEs in a model subject with significant internal reflection. The solid line shows  $|P_{\text{SFOAE}}(f)|$  computed using  $|R_{\text{stapes}}|=0.8$  (with max  $|R|=1$ ). Standing-wave resonances due to multiple internal reflection are clearly visible. For comparison, the dotted line shows the SFOAEs computed for the same *in silico* subject using  $R_{\text{stapes}}=0$ . The two dashed lines show estimates of the direct emission obtained using signal-processing methods described in the text. The gray dashed line shows  $|P_{\text{SFOAE}}^0(f)|$  computed using the continuous wavelet transform (CWT) to perform time-frequency filtering (see also Fig. 10). The black dashed line shows  $|P_{\text{SFOAE}}^0(f)|$  computed using cepstral smoothing and the  $\phi$ FFT. The mean SNR was 25 dB.

introductions to wavelet analysis and its application to click-evoked OAEs are available elsewhere (e.g., Wit et al., 1994; Tognola et al., 1997). Although the property has seldom (if ever) been employed in otoacoustic applications, the CWT is also invertible, allowing the reconstruction of signals from a set of (possibly modified) wavelet coefficients. We exploit that feature in the algorithm developed here.

At delay  $\tau$  and frequency  $f$ , the continuous wavelet transform of the signal  $p(t)$  is defined by the following integral:<sup>10</sup>

$$\Psi_p(\tau, f) = \int p(t) \sqrt{f} \psi^*(f(t - \tau)) dt, \quad (8)$$

where  $\psi(t)$  is the mother wavelet. In our application, the mother wavelet is complex-valued, and the asterisk represents the operation of complex conjugation. Given the transform  $\Psi_p(\tau, f)$ , the signal can be reconstructed using the inverse CWT, defined by

$$p(t) = C_{\psi}^{-1} \iint \Psi_p(\tau, f) \sqrt{f} \psi(f(t - \tau)) d\tau df, \quad (9)$$

where  $C_{\psi}^{-1}$  is a constant that depends on the wavelet. We use a complex Morlet mother wavelet of the form,

$$\psi(t) \sim e^{2\pi i t} e^{-t^2}, \quad (10)$$

a function whose real and imaginary parts are plane waves (or pure tones) localized by a Gaussian window. Other reasonable choices of the mother wavelet—such as those previously used to analyze click-evoked emissions (e.g., Wit et al., 1994; Tognola et al., 1997; Sisto et al., 2007) and whose temporal envelope resembles the frequency response of a low-pass Butterworth filter [e.g., wavelets of the form  $\psi(t) \sim \cos(\omega_0 t)/(1 + |t|^n)$ —yield similar results.

Figure 10 (top) shows a grayscale image of the magnitude of the complex CWT,  $|\Psi_p(\tau, f)|$ , computed for the same simulated SFOAE shown in Fig. 9. The time-domain SFOAE waveform used in the analysis,  $p_{\text{SFOAE}}(t)$ , was computed from  $P_{\text{SFOAE}}(f)$  using the inverse fast Fourier transform (FFT). Long-lasting emission components are visible at multiple frequencies, and are especially prominent in the region about 1 kHz, corresponding to the two tall peaks in Fig. 9. The solid line gives the value of the energy-weighted group delay, as computed from the CWT using the following formula:

$$\tau_{\text{EW}}(f) \equiv \int \tau |\Psi_p(\tau, f)|^2 d\tau / \int |\Psi_p(\tau, f)|^2 d\tau, \quad (11)$$

where the value  $|\Psi_p(\tau, f)|^2$  provides a measure of the SFOAE energy at every location  $(\tau, f)$  in the time-frequency plane. The integration extends over the full extent of the time waveform. Although the image in Fig. 10 is truncated at  $\tau = 35$  ms, the transform was continued until the emission disappeared into the noise, and the computation of  $\tau_{\text{EW}}(f)$  is therefore not biased downward by the display. To ensure that the value of  $\tau_{\text{EW}}(f)$  is not biased upward by noise, whose energy continues at long times and tends to increase the



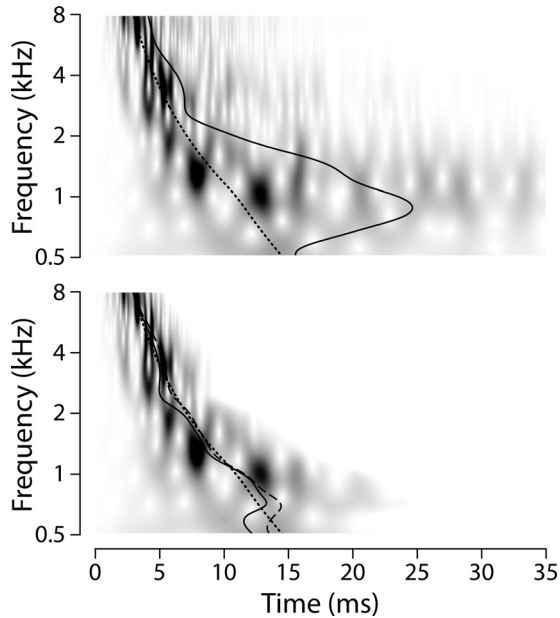


FIG. 10. Continuous wavelet transforms of  $p_{\text{SFOAE}}(t)$  for an *in silico* subject with significant internal reflection ( $|R_{\text{stapes}}| = 0.8$  and  $\max |R| = 1$ ). The grayscale images show the magnitude of the complex CWT versus time and frequency. Figure 9 shows a section of  $|P_{\text{SFOAE}}^0(f)|$  for the same simulated subject. The bottom panel shows an estimate of the CWT of the direct emission, obtained using the procedure described in the text (wavelet coefficients at times  $\tau > \alpha \tau_{\text{SFOAE}}^0$ , with  $\alpha = 1.9$ , have been masked out). In both panels, the solid lines show values of the energy-weighted group delay computed from the corresponding CWT using Eq. (11). The dotted lines show the trend  $\bar{\tau}_{\text{SFOAE}}(f)$  used in the model computations. In the bottom panel, the dashed line shows the group delay for SFOAEs computed using the same model parameters but no internal reflection ( $R_{\text{stapes}} = 0$ ).

energy-weighted group delay, we thresholded the transform  $|\Psi_p(\tau, f)|$ , setting to zero those coefficients with magnitudes less than 6 dB above the corresponding value of the transform of the measured noise. (After resynthesis, this amounts to a wavelet-based de-noising.)

For comparison, the dotted line in Fig. 10 shows the phase gradient delay,  $\bar{\tau}_{\text{SFOAE}}(f)$ , expected for the direct emission  $\hat{P}_{\text{SFOAE}}^0$  based on the underlying model. As anticipated, a multiple internal reflection prolongs the energy-weighted delay. Although  $\tau_{\text{EW}} \geq \bar{\tau}_{\text{SFOAE}}$  at all frequencies, the ratio of the two varies irregularly with frequency in a manner idiosyncratic to each subject. In this example, the largest effect occurs about 1 kHz (i.e., near the sharp spectral peaks evident in Fig. 9), where standing waves increase the group delay by almost a factor of 2. The presence of significant multiple reflection thereby complicates interpretation of the emission delay, rendering it an unreliable guide to the value of model parameters.

Measures of delay more easily associated with the underlying mechanics can be obtained by removing the confounding influence of higher-order reflections from the data. The invertibility of the CWT suggests that this might be accomplished by time-frequency filtering. For example, if all wavelet coefficients except those dominated by the direct emission were identified and set to zero (e.g., by multiplying the transform by an appropriate time-frequency mask), an estimate of  $p_{\text{SFOAE}}^0(t)$ , and hence of  $P_{\text{SFOAE}}^0(f)$ , could be obtained by inverting the masked transform using Eq. (9). Any of the

techniques previously described for estimating delay trends could then be applied to the extracted direct SFOAE.

Figure 10 (bottom) shows the results of applying an appropriate time-frequency filter (mask) to the transform given in Fig. 10 (top). The mask removes the long-latency components and yields a transform that closely resembles the transform of the direct emission computed from the same subject (i.e., using the same model parameters and irregularities, but with  $R_{\text{stapes}} = 0$  to preclude the possibility of multiple internal reflection). The energy-weighted group delay computed from the masked transform (solid line) provides a good estimate of both the group delay of the actual direct emission (dashed line) and of the underlying model trend (dotted line). The gray dashed line in Fig. 9 shows the magnitude of  $P_{\text{SFOAE}}^0(f)$  obtained by synthesizing an estimate of  $p_{\text{SFOAE}}^0(t)$  from the masked wavelet coefficients using the inverse transform. The results provide a good approximation to the emission magnitude computed with  $R_{\text{stapes}} = 0$  (dotted line in Fig. 9).

The mask used to obtain the bottom panel of Fig. 10 by filtering out higher-order reflections was computed using an empirical estimate of the energy-weighted group delay of the direct emission. How was this estimate and its dependence on frequency obtained? We begin by regarding the time evolution of the emission revealed by the CWT in a band centered on frequency  $f$  as a succession of “bursts” corresponding to the terms in the power series in Eq. (7) (Zweig and Shera, 1995; Konrad-Martin and Keefe, 2003). We denote the group delay of the first burst (i.e., of the direct or zeroth-order emission) at frequency  $f$  by  $\tau_{\text{SFOAE}}^0$ . Since the group delay measures the latency of roundtrip energy propagation, we expect the center of energy of the second burst (i.e., of the first-order reflection, proportional to  $RR_{\text{stapes}}$ ) to occur around time  $2\tau_{\text{SFOAE}}^0$ . If we imagine labeling the energy by its association with one burst or the other, then sometime between times  $\tau_{\text{SFOAE}}^0$  and  $2\tau_{\text{SFOAE}}^0$  the energy switches over from being associated primarily with the first burst to being associated primarily with the second. If the bursts had equal amplitudes, the switch-over would occur near  $1.5\tau_{\text{SFOAE}}^0$ . However, because  $|RR_{\text{stapes}}| < 1$ , the first burst is generally the larger of the two, and the switch-over occurs somewhat later (i.e., around the time we denote by  $\alpha\tau_{\text{SFOAE}}^0$ , where the value of the multiplicative constant  $\alpha$  is somewhere in the range  $1.5 < \alpha < 2$ ).

We now define the “partial group delay,”  $\tau_{\text{EW}}(f, t)$ , as the group delay computed based on the energy at frequency  $f$  that returns before time  $t$  as follows:

$$\tau_{\text{EW}}(f, t) \equiv \int_0^t \tau |\Psi_p(\tau, f)|^2 d\tau / \int_0^t |\Psi_p(\tau, f)|^2 d\tau. \quad (12)$$

With this definition, the group delay of the first burst (the direct emission) can be approximated by the partial group delay

$$\tau_{\text{SFOAE}}^0(f) \cong \tau_{\text{EW}}(f, \alpha\tau_{\text{SFOAE}}^0). \quad (13)$$

In other words, at frequency  $f$  we approximate the delay of the direct emission ( $\tau_{\text{SFOAE}}^0$ ) by the group delay of the first

“burst,” defined as the partial group delay of the energy that returns before  $\alpha$  times the delay of the direct emission ( $\alpha\tau_{\text{SFOAE}}^0$ ). Although the description sounds circular because the functional equation for  $\tau_{\text{SFOAE}}^0(f)$  cannot be directly evaluated—the delay  $\tau_{\text{SFOAE}}^0$  appears on both sides of the equals sign—the equation can be solved by iteration. Specifically,

$$\tau_{\text{SFOAE}}^0(f) \cong \lim_{n \rightarrow \infty} \tau^{(n)}(f, \alpha), \quad (14)$$

where

$$\tau^{(n)}(f, \alpha) = \tau_{\text{EW}}(f, \alpha\tau^{(n-1)}) \quad (n = 1, 2, \dots), \quad (15)$$

with  $\tau^{(0)}(f, \alpha) = \tau_{\text{EW}}(f, \infty)$ . In practice, one terminates the iteration when the change between successive estimates (e.g., the fractional difference  $|1 - \tau^{(n-1)}/\tau^{(n)}|$  averaged over frequency) is less than some criterion value.<sup>11</sup>

Given the estimate of  $\tau_{\text{SFOAE}}^0(f)$  obtained in Eq. (14), the mask at frequency  $f$  is just a window that tapers quickly to zero at times greater than  $\alpha\tau_{\text{SFOAE}}^0(f)$ . In our application, we used tenth-order recursive exponential windows (Shera and Zweig, 1993). We note that this procedure for estimating  $P_{\text{SFOAE}}^0(f)$  by resynthesizing it from the masked transform makes no assumption about how the latency of the direct emission varies with frequency; rather, the latency  $\tau_{\text{SFOAE}}^0(f)$  is derived from the data. The procedure does, however, depend on the value of  $\alpha$ , but only weakly. We used  $\alpha = 1.9$  but found that values anywhere in the range 1.6–1.9 gave very similar results, independent of the value of  $RR_{\text{stapes}}$ .<sup>12</sup>

To evaluate the utility of time-frequency prefiltering using the CWT, we performed multiple simulations of the type used to produce Fig. 2, but with  $|R_{\text{stapes}}| = 0.8$ . After removing higher-order reflections (standing waves) from the data using the CWT as in Fig. 10, we used the peak-picking algorithm to extract estimates of the model delay trend  $\bar{N}_{\text{SFOAE}}(f)$ . Figure 11 shows the resulting mean estimation errors for  $\bar{N}_{\text{SFOAE}}(f)$ . When trends are extracted without benefit of CWT prefiltering (lines marked with solid squares), standing waves bias estimates of the underlying delay trend upward; for the parameters used here the mean bias amounts to roughly 20%–30%. Although the averaging across subjects performed here yields a mean bias substantially lower than extremes seen in some subjects (cf. Fig. 10, where the bias approaches 100% near 1 kHz), the bias remains significant. Prefiltering with the CWT reduces the bias almost entirely (black lines with solid diamonds); after filtering, the expected estimation error, averaged across frequency, drops to less than 1%, with the largest errors below 1 kHz.

An ideal filtering method would eliminate standing waves from data sets that have them while leaving data without such components unchanged. To see how closely CWT prefiltering approaches the ideal, we also applied the method to data from the same *in silico* subjects computed using  $R_{\text{stapes}} = 0$ . The results shown in Fig. 11 (gray lines with open diamonds) indicate that prefiltering data that lack standing-wave components lead to a small bias in the negative direction, underestimating the trend by an average of 5% (more at lower frequencies, less at higher frequencies).

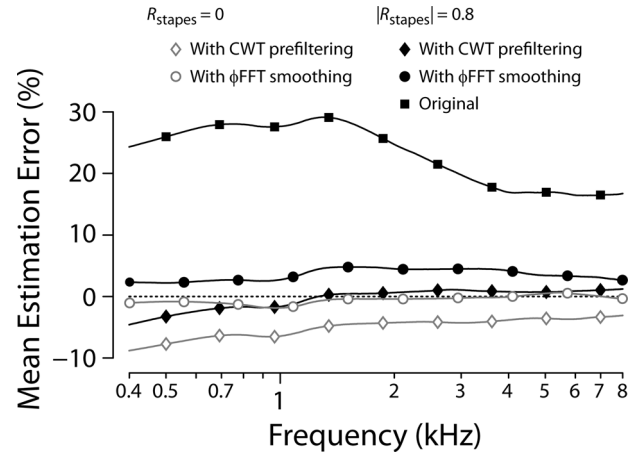


FIG. 11. Mean percent estimation errors for the model trend  $\bar{N}_{\text{SFOAE}}$  versus frequency computed both before (squares) and after (other symbols) mitigating the influence of standing waves using either CWT prefiltering (diamonds) or  $\phi$ FFT smoothing (circles). Black lines marked with closed symbols show results computed using  $|R_{\text{stapes}}| = 0.8$  and max  $|R| = 1$ ; gray lines with open symbols show results computed in the same *in silico* subjects using  $R_{\text{stapes}} = 0$ . Although estimates of  $\bar{N}_{\text{SFOAE}}$  trends are biased upward by the presence of significant intracochlear standing waves (squares), their effects can largely be removed from the data. All group delay trends were estimated using the peak-picking algorithm. Due to the time required to compute and invert the CWTs, the results here are based on a smaller number of *in silico* subjects (750 in 50 groups of 15). The mean SNR in the simulations was 25 dB.

## B. Cepstral smoothing using the $\phi$ FFT

Although time-frequency prefiltering using the CWT yields good results overall, the estimation errors vary significantly across frequency. In particular, the method performs less well below 1 kHz, where deviations from scaling, both in our model and in real data, are largest. In addition, the method is computationally expensive. Although not prohibitive on the smaller data sets used in most studies, the computation time rendered the method more difficult to evaluate on our full complement of 1500 *in silico* subjects. (Indeed, for the results shown in Fig. 11 we used only half that many.) Finally, the method introduces a small negative bias when applied to data without multiple reflections. To address these shortcomings, we developed an alternative procedure that yields better results in a much shorter time. The method is a novel variant of cepstral analysis. Originally introduced for a purpose similar to ours—to separate signals containing echoes (Bogert *et al.*, 1963)—cepstral analysis involves taking the logarithm of the frequency response in order to decompose a product of spectra into a sum. Unlike the more familiar case of time-domain filtering, the oscillatory function to be removed occurs in the frequency response. As a result, cepstral smoothing involves a reversal of the roles usually played by time and frequency.

To explain the method, we begin by noting that Eq. (7) implies that the logarithm of  $\hat{P}_{\text{SFOAE}}$  has the form

$$\log \hat{P}_{\text{SFOAE}} = \log \hat{P}_{\text{SFOAE}}^0 - \log(1 - RR_{\text{stapes}}), \quad (16)$$

representing the sum (superposition) of two components, the first arising from the direct emission and the second, which vanishes when  $R_{\text{stapes}} = 0$ , from multiple internal reflection

within the cochlea. Here we exploit the fact that the spectral signatures of the two additive components are typically quite different. As illustrated in Fig. 9, for example, a nonzero value of  $R_{\text{stapes}}$  can introduce additional “spectral ripples” in the log magnitude (and phase) of  $P_{\text{SFOAE}}$ . But unlike the notches and other macrostructure evident in  $R$  itself, the ripples due to multiple internal reflection have a quasiperiodic oscillatory form whose period is determined largely by the phase gradient of  $R$  (e.g., the ripple period corresponds to one cycle of phase rotation of  $RR_{\text{stapes}}$ ). Because of their different frequency dependencies, the two terms in Eq. (16) can be separated by filtering in the corresponding Fourier domain (e.g., by windowing in the time domain) using the variant of cepstral analysis described below.

The application of cepstral or any sort of inverse fast Fourier transform (IFFT) analysis to SFOAEs is complicated by cochlear dispersion. Because of dispersion, the overall slope of the phase of  $R$  varies with frequency (i.e., the phase versus frequency function is curved); consequently, emissions at different frequencies are delayed by different amounts, and the period of the spectral interference ripples created by multiple reflection varies with frequency. By smearing things out across time and frequency, cochlear dispersion makes it more difficult to separate the direct emission component from the succeeding echoes. As explored in Sec. VA, time-frequency analysis using wavelets provides one way of addressing this problem. Alternatively, one can work with a transformed frequency coordinate. Ideally, the transformation would compensate for the dispersion by ensuring that the “transformed group delay” (i.e., the slope of the emission phase computed with respect to the new frequency variable) remains constant. Previous studies using IFFT methods or spectral smoothing to separate OAE components have used logarithmic or power-law transformations to approximate this ideal (Knight and Kemp, 2001; Kalluri and Shera, 2001). One limitation of these previous transformations is that the assumed frequency dependence of the phase is not generally valid throughout the cochlea.

Here we circumvent this limitation and tailor the method to each subject individually by using a transformation derived directly from the emission phase itself. Specifically, we compute Fourier transforms with respect to the variable

$$\phi \equiv -\widetilde{\arg}P_{\text{SFOAE}}(f)/2\pi, \quad (17)$$

where  $\arg$  extracts the *unwrapped* phase and the diacritical tilde denotes subsequent smoothing. The smoothed phase  $\phi$  is an estimate of the secular variation of the unwrapped phase from which much of the phase rippling pattern has been ironed out. Smoothing the unwrapped phase is necessary here to render the transformation monotonic; the minus sign then guarantees that  $\phi$  increases with  $f$ . Later in the analysis, the smoothed phase  $\phi$  is removed from  $\arg P_{\text{SFOAE}}$  [cf. Eq. (18)] to obtain the phase ripples themselves. How the smoothing is performed is not critical; we used Savitzky–Golay filters, but smoothing splines or other methods would work just as well. We call the FFT computed with respect to this unconventional frequency variable the  $\phi$ FFT.<sup>13</sup>

We denote the value of  $P_{\text{SFOAE}}(f)$  with internal reflections removed by  $P_{\text{SFOAE}}^0(f)$ , where the “0” represents the effective value of  $R_{\text{stapes}}$ . We estimate  $P_{\text{SFOAE}}^0(f)$  by cepstral smoothing using the following formula:

$$\log P_{\text{SFOAE}}^0 \cong F^{-1}\{WF\{\log P_{\text{SFOAE}} + 2\pi i\phi\}\} - 2\pi i\phi, \quad (18)$$

where  $F\{\cdot\}$  represents the operation of Fourier transformation with respect to  $\phi$ ,<sup>14</sup>  $F^{-1}\{\cdot\}$  is the inverse transform with respect to  $\eta$  (our name for the variable conjugate to  $\phi$ ), and  $W(\eta; \eta_c)$  represents the application of a window with cutoff at  $\eta_c$ . In the language of cepstral analysis,  $\eta$  is the queffrequency or “time” variable; application of the window  $W$  thus smooths the response by eliminating long “latency” components in the complex cepstrum.

To perform the  $\phi$ FFT smoothing numerically, we resampled values of  $\log P_{\text{SFOAE}}(f)$  at frequencies corresponding to equal intervals of  $\phi(f)$  using cubic spline interpolation.<sup>15</sup> Because measurements are only available over a finite frequency range, the cepstral smoothing operation is complicated by end effects and the assumption of periodic boundary conditions employed in the Fourier analysis. To mitigate these effects by removing the secular variation of the phase, we added  $2\pi i\phi$  to  $\log P_{\text{SFOAE}}$ , a function later subtracted back in as indicated in Eq. (18). In addition, the analyzed frequency range was chosen to include an integer number of cycles of  $\phi$  (and thus an approximately integral number of spectral ripples), and linear ramps were subtracted (and restored after smoothing) to render the arguments to the Fourier transforms periodic. So that the window would remove queffrequencies associated with standing-wave ripples, we chose the value of the window cutoff based on the total number of cycles traversed by the phase  $\phi$ . Denoting this number of cycles by  $N_\phi$ , we used  $\eta_c = 0.9N_\phi$ . The window was implemented using recursive exponential filters (Shera and Zweig, 1993; Kalluri and Shera, 2001). Finally, estimates of  $P_{\text{SFOAE}}^0$  were discarded at each end over frequency intervals equal to the approximate bandwidth of the smoothing function (to allow for this, the frequency range of the simulations was extended by an octave in both directions).

To illustrate the procedure, Fig. 12 shows mean  $\phi$ FFT cepstra computed from  $\log|P_{\text{SFOAE}}|$  for subjects with and without significant standing-wave components in their emission spectra. Cepstra for the two groups differ principally at queffrequencies  $\eta \geq N_\phi$  (i.e., at values greater than unity along the abscissa). Components at these high queffrequencies (long latencies) are significantly enhanced by multiple internal reflection. Although the period of the standing-wave peaks and valleys varies with frequency, computing the transform with respect to  $\phi$ , rather than with respect to  $f$ , compensates for cochlear dispersion and produces the sharp peak in the transform near  $\eta \cong N_\phi$ . By removing components at queffrequencies greater than  $N_\phi$  using the window  $W(\eta, \eta_c)$ ,  $\phi$ FFT smoothing largely eliminates standing-wave components from the response. The black dashed line in Fig. 9 shows that the estimate of  $P_{\text{SFOAE}}^0(f)$  obtained by  $\phi$ FFT smoothing provides an excellent approximation to the emission magnitude computed with  $R_{\text{stapes}} = 0$ .

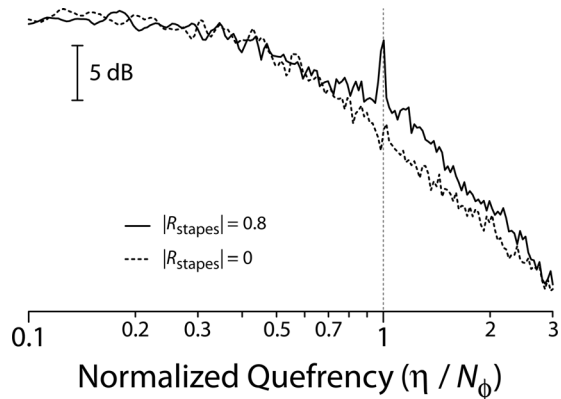


FIG. 12. Mean  $\phi$ FFT cepstra in subjects with and without intracochlear standing waves. The solid line shows  $|F\{\text{Re}\{\log P_{\text{SFOAE}}\}\}|$  versus normalized quefreny averaged over 50 *in silico* subjects with  $|R_{\text{stapes}}| = 0.8$  (and  $\max|R| = 1$ ). Since the number of potential standing-wave maxima in  $|P_{\text{SFOAE}}(f)|$  varies from subject to subject, the quefreny  $\eta$  was normalized by  $N_\phi$  before averaging. The sharp peak near unity along the abscissa (vertical gray dashed line) arises from spectral components in  $P_{\text{SFOAE}}(f)$  that originate via multiple internal reflection.  $\phi$ FFT smoothing removes the peak and other high quefreny components by windowing the cepstrum. The dotted line shows the corresponding mean cepstrum computed in the same subjects with  $R_{\text{stapes}} = 0$ .  $\phi$ FFT cepstral magnitudes were averaged across subjects to reduce the variance and render the systematic differences between those with and without significant standing wave components more salient. The mean SNR was 25 dB.

Estimation errors obtained after using  $\phi$ FFT smoothing to remove standing-wave components from the data are shown in Fig. 11 (lines with solid circles) for the same subjects used to evaluate CWT prefiltering. Although the expected estimation errors are slightly larger than those obtained using CWT prefiltering at high frequencies, where they average  $\sim 3\%$ , the error is more uniform across frequency. In addition, smoothing with the  $\phi$ FFT has a rather benign effect on data sets in which higher-order reflections are small or absent (open circles). To understand this, note that the high quefrencies removed by the windowing operation come from the parts of the SFOAE spectrum that change rapidly with frequency. In data sets without strong standing-wave components, these tend to be the regions around notches, which are precisely the regions that are discounted by analysis methods such as peak-picking. Thus,  $\phi$ FFT smoothing applied to data free of significant standing-wave ripples has only a minor effect on phase-gradient delay trends.

## VI. DISCUSSION

We performed *in silico* experiments to evaluate both previously and newly proposed strategies for extracting phase-gradient delay trends from reflection-source OAE data. Our results show that despite measurement noise and the intrinsic irregularity of the emission spectra, robust and minimally biased estimates of OAE phase-gradient delay trends can be reliably extracted using most of the analysis methods described here. Four of the six methods we evaluated (namely, peak-picking, energy-weighting, SNR-based data exclusion, and nothing special) reliably extracted the underlying model trend (i.e., with a bias less than a few per-

cent) when employed at sufficiently large SNRs (i.e.,  $\text{SNR} \geq 10$  dB). Peak-picking and energy-weighting produced estimates with the smallest uncertainties and were the top overall performers. The remaining two methods (all-pass factorization and phase smoothing) both yield significantly biased estimates of the delay trend and are not recommended.

### A. Role of the model

To test and validate the various analysis methods, we simulated reflection-source OAEs using a phenomenological model of the emission process. Although the model inherits its ability to generate realistic SFOAEs from coherent-reflection theory, the same modeling framework need not be used to interpret delay trends extracted from measurements in actual ears. For example, although coherent-reflection theory provides a ready physical interpretation of the phase-gradient delay in terms of the wavelength of the traveling wave in the region of OAE generation, the validity of our conclusions about methods of data analysis do not depend on this interpretation. For our purposes here, the only aspects of the model that really matter are that the model provides a known, benchmark form of  $\bar{N}_{\text{SFOAE}}(f)$  for computing the estimation error and that the simulations produce realistic SFOAE data as fodder for subsequent signal processing.

The parameters of the phenomenological model were chosen to simulate the measured properties of human SFOAEs, in particular, the segmented, approximate power-law form of  $N_{\text{SFOAE}}(f)$  derived from the frequency dependence of the SFOAE phase (Shera and Guinan, 2003). Although the variation of  $N_{\text{SFOAE}}(f)$  with frequency in humans remains the subject of some debate (e.g., Schairer *et al.*, 2006; Shera *et al.*, 2010; Bentsen *et al.*, 2011)—indeed, the hope of informing this debate provided a principal motivation for the current study—our conclusions about the reliability of the various analysis methods are not especially sensitive to the assumed frequency dependence. In numerous control simulations not detailed here, we found qualitatively similar results using other forms for  $\bar{N}_{\text{SFOAE}}(f)$ , such as the pure “scaling” model in which  $\bar{N}_{\text{SFOAE}}$  is assumed constant and  $\bar{\tau}_{\text{SFOAE}}(f)$  varies as  $1/f$ . As discussed in the following, we also sought to identify the principles underlying the relative success or failure of the different methods, and found that, with one exception, none of them hinges on the assumed form of  $\bar{N}_{\text{SFOAE}}(f)$ . The exception is the phase-smoothing method, which introduces a bias whose variation across frequency depends on the global curvature of the emission phase, and thus on  $\bar{N}_{\text{SFOAE}}(f)$ .

Although our simulations were implemented in the frequency domain as an explicit model of human SFOAEs, we expect our results also apply to phase-gradient delays obtained from reflection-source OAEs measured with other paradigms (e.g., by measuring click-evoked emissions or by unmixing distortion-product OAEs to extract the “reflection” component) and in other species (e.g., chinchillas, lizards). Although the optimal analysis parameters will no doubt require adjustment on a case-by-case basis, the methods described here are potentially of wide utility.

## B. Simplifications in the model

Our focus here has been on devising and evaluating strategies for mitigating the influence on estimated phase-gradient delay trends of the large and irregular variations across frequency arising from reflection-source OAE spectral macrostructure. Although the model captures the inherent irregularity of the delay, as well as contributions from Gaussian measurement noise,<sup>16</sup> OAEs recorded from real ears contain other sources of variability not included in the model. For example, different ears, in addition to having different micromechanical irregularity patterns, no doubt present with a range of underlying cochlear tuning and delay characteristics. Restated in terms of model parameters, the real-life analogs of the functions  $\bar{N}_{\text{SFOAE}}(f)$  and  $Q_{\text{ERB}}(\text{CF})$  presumably vary somewhat from subject to subject, even among those with “normal” hearing. Although the model could readily be extended to incorporate this added realism, doing so seems unlikely to alter our conclusions about the relative merits of the proposed analysis strategies. [It would, of course, affect such things as the standard deviation of the prediction error (Fig. 6), which would be increased by the additional variability in the population.]

By setting  $G_{\text{ME}} = 1$ , we elected, in effect, to ignore emission features arising from roundtrip middle-ear transmission. Although the middle ear plays a central role in determining overall emission levels, the forward- and reverse-transfer functions generally vary only slowly with frequency compared to SFOAE phase or spectral macrostructure (e.g., Puria, 2003) and are thus unlikely to be major determinants of emission delay or its variability.<sup>17</sup> We therefore expect that extending the model to include a more realistic form of  $G_{\text{ME}}(f)$  and its possible variation from subject to subject will leave our basic conclusions unchanged. In cases where middle-ear mechanics do contribute substantially to SFOAE delays—the bulla resonance in the tiger may be an example (Bergevin *et al.*, 2012b)—we expect an impact not so much on the choice of signal processing strategy used to extract the delays as on the subsequent interpretation of those delays in terms of cochlear and middle-ear mechanics.

## C. Effects of multiple internal reflection

In addition to evaluating strategies for extracting delay trends, we used our simulations to demonstrate that standing waves caused by multiple internal reflection can complicate the interpretation and application of OAE delays. For example, by increasing emission delays through mechanisms not directly related to mechanical tuning, standing waves can bias estimates of tuning sharpness and its variation with stimulus intensity obtained from OAE phase-gradient delays. Multiple internal reflection is strongest when the product of the cochlear and stapes reflectances,  $RR_{\text{stapes}}$ , is close to one, a condition most likely to occur at stimulus levels near threshold in strongly emitting ears. Our simulations demonstrate both that the bias in individual ears can be large near standing-wave frequencies (cf. Fig. 10) and that averaging across subjects appears to reduce the bias in the delay trend substantially. Consistent with this latter result, the one study

that compared phase-gradient delays in ears with and without SOAEs found no significant differences in the pooled trends (Bergevin *et al.*, 2012a). Although those who wish to measure emission latencies at low levels or in special populations should be aware of the potential complications, we suspect that bias introduced by multiple reflection is likely insignificant in most published measurements of reflection-source OAE latency trends—most of which either pooled data across individuals with a range of emission strengths, excluded data near spontaneous emissions, and/or made measurements at stimulus levels where the appearance and spacing of the OAE magnitude maxima suggests that  $|RR_{\text{stapes}}|$  was much less than one. We therefore expect that the methods proposed here will be most useful in addressing standing-wave effects in individual ears.

To separate out the effects of multiple reflection, we developed and evaluated two signal-processing procedures—prefiltering with continuous wavelet transforms and  $\phi$ FFT smoothing—to isolate the direct emission from its later reflections and thereby eliminate any standing-wave components that may be present in the measurements. (Although we here emphasize the application of CWT prefiltering and  $\phi$ FFT smoothing for eliminating standing-wave components in order to obtain the direct emission, the same methods can of course be used to extract and study the standing-wave component itself.) Although both methods substantially reduce the estimation bias for the trend,  $\phi$ FFT smoothing provides more uniform results across frequency and has a more benign effect on data free of standing-wave ripples. In addition,  $\phi$ FFT smoothing is both easier to implement and more computationally efficient. (Unlike CWT prefiltering, however, the  $\phi$ FFT method lacks the intuitive appeal provided by colorful images that make visible the multiple reflections as they echo across time and frequency.) Because the CWT method relies on analysis and resynthesis using a bank of constant- $Q$  filters whose properties are not optimally matched to those of the emissions being analyzed—which originate from a system in which tuning and delay violate the scaling assumption of constant- $Q$ —the method introduces a frequency-dependent bias whose effects are especially apparent below 1 kHz, where violations of scaling are most pronounced.

## D. Peak-picking and related methods

The two top performing methods (peak-picking and energy-weighting) yield close to statistically unbiased estimates of the underlying model trend (Fig. 2), have the smallest standard deviations (Fig. 3), and produce trends with the least uncertainty. Both methods emphasize data near peaks in SFOAE magnitude, where the phase gradient is generally closest to the model trend (Fig. 6). Of these two methods, energy-weighting is the easiest to implement, requiring nothing more than routines for performing weighted loess smoothing. The method of excluding data with poor SNR performs nearly as well, at least when the criterion is chosen to mimic the top two strategies by excluding data at frequencies away from magnitude peaks.

Although manipulations such as energy-weighting and SNR-based data exclusion are usually justified by the

perceived need to discount noisy data, the secret to their success lies primarily in their ability to mimic the peak-picking procedure by emphasizing data near peaks, where delays are closest to the trend. For example, the SNR-based exclusion method only matches the performance of peak-picking when the criterion is chosen to pass the peaks and exclude the rest. Failing to exclude data near magnitude dips, even if the SNR is everywhere exceptional, reduces the ability to recover the underlying trend.

An important virtue of peak-picking and related strategies is that they render estimates of phase-gradient delay trends insensitive to the most common form of phase-unwrapping error. At least in data without significant standing-wave components, these errors occur mostly near magnitude dips, where the SNR is usually poor and the phase tends to change most rapidly, making it difficult to correctly resolve unwrapping ambiguities. As a consequence of their concentration near magnitude dips, unwrapping errors remain largely invisible to strategies that focus on phase gradients near peaks.

The easiest-to-implement method of doing “nothing special” (i.e., simply fitting unweighted robust loess curves to all the available data) worked surprisingly well. We attribute this to the power of locally linear regression, especially *robust* loess smoothing, to extract trends from “noisy” data. In separate experiments, we found that robust fitting, in which an initial trend is iteratively improved after deemphasizing data with large residuals (Cleveland, 1993, Sec. 3.4), noticeably improved the performance of all methods.

Our focus in this paper has been on pooling data across subjects to estimate group or species trends. However, by reducing the scatter inherent in reflection-source OAE delay data, peak-picking and related methods also hold promise for obtaining more reliable estimates of phase-gradient delays, and thus perhaps estimates of cochlear tuning, in individuals. (Some evidence for the reliability of the procedures applied to individual subjects appears in Fig. 5 at  $n = 1$  along the abscissa.) Removing the effects of multiple internal reflection is likely to prove critical to the meaningful estimation of OAE delays in individual ears.

Finally, it goes (almost) without saying that although the methods reviewed here can improve the reliability of delay trends extracted from OAE measurements, there can be little substitute for densely sampled data with good signal-to-noise ratios. Even the most effective algorithms, we found, require that the peaks be readily distinguishable from the noise and that phases near those peaks be correctly unwrapped. Peak-picking and related strategies are no exceptions to the rule of “garbage in, garbage out.”

## E. Why peak-picking works

The success of the peak-picking method, when applied to data without significant standing-wave components, can be understood heuristically by considering the complementary question of why one should avoid spectral notches and dips. Notches in emission magnitude and their associated phase ripples arise when two or more emission “components” combine by vector addition and nearly cancel one another. The multiple combining components need not arise from spatially dis-

tinct regions of the cochlea (e.g., from the tip and tail regions of the traveling wave). For example, notches in SFOAE magnitude can occur when the wavelets scattered by irregularities located in the apical and basal halves of the traveling-wave peak region happen to have similar amplitudes, but nearly opposite phases. In this case, the emission components arise from the same general region, albeit from different locations within the peak. At frequencies near notches, emission magnitudes and phases—and hence emission phase-gradient delays—are precariously sensitive to changes in the individual components. As a result, small changes (e.g., in the irregularities contributing to the emission because of shifts in stimulus frequency or level) can produce significant but largely uninformative effects. For this reason, phase gradients near notches are often dominated by local, idiosyncratic mechanisms and can be unreliable guides to the overall emission delay. Conversely, emission phase and delay near spectral maxima, which occur when components with similar phases combine constructively, are not nearly so sensitive to small changes in the individual components. For the same reason, it is also near emission maxima that the spectral magnitude tends to vary most slowly with frequency, a necessary condition for interpreting phase gradients as actual physical time delays (Papoulis, 1962, Sec. 7–5).

## 1. A two-component model for peaks and dips

These heuristic statements about why peak-picking works can be made more precise with the help of a simple two-component model for SFOAE peaks and dips. For this purpose, we represent local SFOAE magnitude and phase by a function of the following form:

$$F(f) \sim [1 + Ae^{i\theta}] e^{-2\pi i f \tau_1}, \quad (19)$$

consisting of a general sum of two components, with relative amplitude  $A > 0$  and phase  $\theta$ , multiplied by a phasor that captures the overall emission delay ( $\tau_1 > 0$ ). We imagine that the relative phase of the two components,  $\theta(f)$ , varies slowly with frequency (i.e.,  $|\tau_2|/\tau_1 \ll 1$  where  $2\pi\tau_2 \equiv -d\theta/df$ ). The interaction of the two components produces a peak in  $\log|F|$  (i.e., in the SFOAE level) when the components are in phase ( $\theta = 0$ ) and a dip when they are out of phase ( $\theta = \pm\pi$ ). The depth of the dip depends on  $A$ ; when  $A$  is close to 1 the two components nearly cancel and the dip becomes a deep notch. Fixing the relative amplitude of the components and computing the phase-gradient delay yields

$$\tau = \tau_1 + \tau_2 \frac{A^2 + A\cos\theta}{1 + A^2 + 2A\cos\theta}, \quad (20)$$

where the first term ( $\tau_1$ ) is the overall delay of interest and the second term represents the contribution to the phase gradient arising from interference between the components. The interference term is largest near a dip, where the phase-gradient delay becomes

$$\tau_{\text{dip}} = \tau_1 - \tau_2 \frac{A}{1 - A}. \quad (21)$$

Because of the possible zero in the denominator, the phase-gradient delay can vary wildly near a dip, even when  $|\tau_2| \ll \tau_1$ . Letting  $A = 1 - \epsilon$  yields  $\tau_{\text{dip}} \cong \tau_1 - \tau_2/\epsilon$  near a notch ( $|\epsilon| \ll 1$ ). Thus, depending on the relative amplitude of the components,  $\tau_{\text{dip}}$  can easily be much larger or much smaller than  $\tau_1$ . Indeed,  $\tau_{\text{dip}}$  can even become negative. Conversely, the interference contribution to the phase-gradient delay is smallest near a peak, where

$$\tau_{\text{peak}} = \tau_1 + \tau_2 \frac{A}{1 + A}. \quad (22)$$

In this case, the phase-gradient delay can be either longer or shorter than  $\tau_1$ , depending on the sign of  $\tau_2$ , but the total deviation from  $\tau_1$  is limited. When the two components have similar amplitudes,  $\tau_{\text{peak}} \cong \tau_1 + \frac{1}{2}\tau_2$ , which is close to  $\tau_1$  when  $|\tau_2| \ll \tau_1$ . Our analysis thus supports the computational results illustrated in Fig. 6. In particular, the phase-gradient delay generally appears most representative of the underlying trend at frequencies corresponding to emission maxima.

## F. Why all-pass factorization fails

The all-pass factorization method relies on the presumption that the “notchy” behavior seen in SFOAE spectra is minimum-phase—or at least sufficiently close to being minimum-phase to be useful—and can therefore be divided out to obtain an all-pass component consisting of a more smoothly varying delay. As illustrated in Fig. 8, however, the delay of the all-pass component is not uniformly smoother than that of the total SFOAE, even at high SNR—sometimes it is, sometimes it is not. The issue is not principally one of measurement noise; similar results are obtained even when the noise is zero. Instead, the failure of the method implies that SFOAE macrostructure is not, in fact, approximately minimum-phase. Equivalently, the all-pass or delay component of reflection-source OAEs [i.e., the delay denoted  $\tau_{\text{AP}}$  in the text following Eq. (6)] is inherently irregular (cf. Kalluri and Shera, 2012).

The absence of minimum-phase behavior in SFOAE macrostructure can readily be understood using the two-component model for SFOAE peaks and dips (see Sec. VI E 1). In that context, the presumption underlying the all-pass factorization is that the sum of the two components [i.e.,  $1 + Ae^{i\theta}$  in Eq. (19)] can be identified with  $P_{\text{MP}}(f)$  in Eq. (6). The complex phasor representing the overall delay [i.e.,  $e^{-2\pi if\tau_1}$  in Eq. (19)] is then the all-pass component. Whether a function is minimum-phase or not depends on the locations of its poles and zeros in the complex frequency plane (Papoulis, 1962, Sec. 10–3). In general, there is no guarantee that a sum remains minimum-phase, even if both components qualify individually. In this case, the locations of the zeros of the sum  $1 + Ae^{i\theta}$  depend sensitively on the complex ratio of the two components (i.e., on their relative amplitude and phase), which evidently varies in an irregular way from notch to notch; consequently the sum is not generally minimum-phase. Indeed, explicit computation of the zeros of  $1 + Ae^{i\theta}$  for a simple form of  $\theta(f)$  shows that they often violate the

criteria for minimum-phase behavior. When the phase is a pure delay of the form  $\theta(f) = -2\pi f\tau_2$ , the zeroes occur when  $Ae^{-2\pi if\tau_2} = -1$ ; that is, at those (complex) frequencies  $f_z$  satisfying

$$2\pi f_z \tau_2 = \left(n + \frac{1}{2}\right) - i \log A \quad (n = 0, \pm 1, \dots). \quad (23)$$

Whenever  $\tau_2$  and  $\log A$  have the same sign (i.e., if  $\tau_2 > 0$  and  $A > 1$ , or  $\tau_2 < 0$  and  $A < 1$ ), then  $\text{Im} f_z < 0$ , the zeroes lie in the lower half of the complex  $f$ -plane, and the function is not minimum-phase. Although minimum-phase behavior is a *global* property that depends on the locations of *all* the poles and zeros, we conjecture that the all-pass factorization method generally smooths well at frequencies near zeroes in the upper half-plane ( $\text{Im} f_z > 0$ ), while having the opposite effect (i.e., increasing the local curvature, or anti-smoothing) near zeroes in the lower half-plane, thereby accounting for the variability in performance evidence in Fig. 8. Although the all-pass factorization method seems a lovely idea *a priori*, SFOAE macrostructure—both *in silico* and, we suspect, *in vivo*—is evidently not minimum-phase (or close to it), and the method therefore fails to produce overall useful smoothing.

## G. Why phase smoothing introduces bias

By overestimating the delay  $\bar{N}_{\text{SFOAE}}$  at low frequencies and underestimating it at high frequencies (Fig. 2), smoothing the phase produces a frequency-dependent bias. This pattern of bias results from the generally downward curvature of the SFOAE phase (see Fig. 1) when plotted versus  $\log f$ , the abscissa used to calculate the smoothing splines.<sup>18</sup> In the limiting case of maximal smoothing ( $s = 1$ ), SFOAE phase curves similar to those shown in Fig. 1 are transformed into straight lines with the same rise and run. However, because of the overall downward curvature of the phase, the slope of these best-fitting lines is much greater than the actual phase slope at low frequencies and much smaller than the actual phase slope at high frequencies. Although the deviations are less extreme when smoothing with smaller values of  $s$ , the same overall pattern of bias remains.

Another problem with phase smoothing, at least when implemented with conventional smoothing splines, is that the minimization involves the total integrated phase curvature. Because of the global optimization, delay estimates at widely different frequencies become linked. Consequently, the SNR or overall phase curvature at high-frequencies influences the amount of smoothing performed at low frequencies, and vice versa, even though the emissions themselves are physically independent. Although we did not attempt to salvage the method by exploring the possibility, allowing the smoothing parameter  $s$  to vary with frequency might help to alleviate these problems.

Although smoothing the phase introduces bias and appears undesirable when estimating phase-gradient delays, smoothing the *delay* itself (e.g., using loess smoothing) proves extremely effective at ironing out subject-dependent

bumps and wiggles to extract reliable and meaningful delay trends. The difference in outcomes can be traced largely to *when* the smoothing is applied: By smoothing the delay rather than the phase, we postpone the smoothing operation, and its potential for distorting the data to produce damaging downstream consequences, until the very end of the process.

Our results corroborate previous reports that smoothing the phase prior to estimating the derivative yields delay trends that differ systematically from those obtained without smoothing (Schairer *et al.*, 2006; Bentsen *et al.*, 2011). The results help resolve differences in the literature about the value and frequency dependence of SFOAE delay and their consequences for noninvasive estimates of human cochlear tuning. In particular, the bias introduced by phase smoothing suggests that trends obtained without smoothing are almost certainly more characteristic of the underlying mechanics and tuning of the cochlea.

## ACKNOWLEDGMENTS

We thank Carolina Abdala, John Guinan, Radha Kalluri, Sarah Verhulst, and the anonymous reviewers for helpful comments on the manuscript. Work supported by Grant No. R01 DC003687 from the NIDCD, National Institutes of Health.

<sup>1</sup>Although the subject is beyond the scope of this paper, the coherent-reflection model indicates that information about cochlear mechanisms can be extracted not only from overall trends but also from details of the frequency fluctuations themselves (e.g., from the pattern of the spacings between amplitude notches).

<sup>2</sup>Here and throughout, we use the term “reliable” both in its dictionary sense (“consistently good in quality or performance; able to be trusted”) and as a shorthand for the more technical “without appreciable bias or uncertainty.”

<sup>3</sup>The phase-gradient delay,  $\tau(f)$ , corresponding to the phase  $\theta(f)$  is defined as  $2\pi\tau = -d\theta/df$ .

<sup>4</sup>The factor of 2 multiplying  $N_W(\text{CF})$  in Eq. (5) comes from the square in Eq. (3) for  $R(f)$ . Physically, the factor of 2 arises because in conventional cochlear models the roundtrip (emission) delay is approximately twice the forward delay of the traveling pressure-difference wave (Shera *et al.*, 2008).

<sup>5</sup>The “bend” in the  $N_{\text{SFOAE}}(f)$  curve appears to be associated with the emergence of an additional, short-latency SFOAE component in the apical region of the cochlea (Shera and Guinan, 2003; Siegel *et al.*, 2005; Shera *et al.*, 2008). Although the origin of the short-latency component remains unclear (Shera *et al.*, 2008), we have incorporated the bend into our phenomenological model to produce more realistic emission delays.

<sup>6</sup>Although we had no need to employ it here, our implementation of the loess smoothing algorithm allows the value of  $\alpha_{\text{loess}}$  to vary with frequency (i.e., with the value of the independent variable). In the current application, this feature allows one to specify that the fitting window span a constant number of octaves, regardless of the local sampling density. Imagine, for example, that we had chosen to sample the data using linear frequency spacing rather than logarithmic. If the fits were performed after log-transforming the frequency, the constant value of  $\alpha_{\text{loess}}$  used in standard loess would then produce a fitting window that spanned a larger (possibly much larger) number of octaves at low frequencies than at high. Our implementation allows one to guarantee that the fit be equally “local”—however one chooses to define that—at all frequencies, independent of the choice of sampling. MATLAB software for computing loess curves and their confidence intervals is available from the authors.

<sup>7</sup>In our simulations, the noise floor and the mean SNR were approximately constant across frequency (Fig. 1). Consequently, we applied the same SNR-based exclusion criteria (e.g., 15 dB) at all frequencies. When ana-

lyzing real measurements, however, it may be desirable, especially when mimicking the peak-picking strategy, to vary the criterion with frequency to compensate for frequency dependence of the noise floor.

<sup>8</sup>MATLAB’s `csaps` function uses a parameter,  $p$ , defined as  $p \equiv 1 - s$ .

<sup>9</sup>We computed the minimum-phase component using the Hilbert transform, as implemented in MATLAB’s `recepss` function. `recepss` requires an estimate of the system’s impulse response, a time waveform equivalent in this case to the model’s click-evoked emission. We obtained the necessary time waveform from the SFOAE using the inverse Fourier transform. Technically, this required interpolating our SFOAE measurements to linear frequency spacing; however, skipping this step and simply inverse transforming the log-spaced measurements had no particular effect, either beneficial or adverse, on the results.

<sup>10</sup>Note that with the substitution  $\sqrt{f}\psi(f) \rightarrow e^{2\pi i f t}$  in Eq. (8),  $\Psi_p(0, f)$  becomes the Fourier transform of  $p(t)$ .

<sup>11</sup>We used a convergence criterion of 1%, but found that the precise value makes little difference. In all but a handful of recalcitrant cases, convergence was achieved in fewer than five to ten iterations.

<sup>12</sup>When computing estimates of  $\tau_{\text{SFOAE}}^0(f)$ , we used a constant value of  $\alpha$  for simplicity. A better, albeit more computationally intensive procedure would be to determine the optimal value of  $\alpha$  adaptively. For example, at each analysis frequency in every subject one could compute the latency  $\tau_{\text{SFOAE}}^0(f, \alpha)$  as a function of  $\alpha$  over some reasonable range (e.g.,  $1.3 < \alpha < 2.3$ ) and identify the value of  $\alpha$  where the dependence on  $\alpha$  is weakest. In other words, one could define the optimal value of  $\alpha$ , for that frequency in that subject, as the value which minimizes the derivative  $\partial \tau_{\text{SFOAE}}^0(f, \alpha) / \partial \alpha$ .

<sup>13</sup>The  $\phi$ FFT is distinguished by both spelling and pronunciation from the slang term “phiff,” which the Urban Dictionary defines as “a word that brain dead matter comes up with when it is unable to produce a more intelligent sound.”

<sup>14</sup>The operation of Fourier transformation with respect to  $\phi$  is defined as  $F\{X(\phi)\}(\eta) = \int X(\phi) e^{-2\pi i \phi \eta} d\phi$ . When applied to a function of frequency,  $G(f)$ , the function  $X(\phi)$  is taken to be  $X(\phi) = X[\phi(f)] \equiv G(f)$ . When  $X(\phi)$  is complex, as in Eq. (18), we have found that computing the transforms of its real and imaginary parts separately often proves convenient. Potential problems with branch cuts in computing  $\log P_{\text{SFOAE}}(f)$  in Eq. (18) can be circumvented by defining the complex logarithm as  $\log z(f) = \log |z(f)| + i\theta(f)$ , where  $\theta(f)$  is the *unwrapped* phase of  $z(f)$ .

<sup>15</sup>Interpolation was necessary when resampling the data at equal intervals of  $\phi(f)$  because we assumed, for consistency with previous measurements, that SFOAE values were only available at a predetermined set of discrete measurement frequencies (see Sec. II). However, we recommend measuring reflection-source OAEs using swept tones (e.g., Choi *et al.*, 2008; Long *et al.*, 2008; Bennett and Özdamar, 2010; Kalluri and Shera, 2012), in which case no interpolation is necessary; the swept responses can simply be reanalyzed to extract SFOAE values at the desired frequencies.

<sup>16</sup>By modeling the measurement noise as Gaussian, we are assuming, in effect, that any non-Gaussian noise contributions to measured SFOAEs (e.g., that horrible coughing fit in the middle of the measurement session) have been successfully removed by artifact-rejection algorithms.

<sup>17</sup>In chinchilla, for example, middle-ear delay appears negligible compared to traveling-wave or otoacoustic delay (e.g., Ruggero *et al.*, 1990; Songer and Rosowski, 2007).

<sup>18</sup>If the smoothing is performed on a linear frequency axis, the phase curvature changes from convex down to convex up, and the pattern of bias across frequency is reversed.

- Bennett, C. L. and Özdamar, Ö. (2010). “Swept-tone transient-evoked otoacoustic emissions,” *J. Acoust. Soc. Am.* **128**, 1833–1844.
- Bentsen, T., Harte, J. M., and Dau, T. (2011). “Human cochlear tuning estimates from stimulus-frequency otoacoustic emissions,” *J. Acoust. Soc. Am.* **129**, 3797–3807.
- Bergevin, C. (2011). “Comparison of otoacoustic emissions within gecko subfamilies: Morphological implications for auditory function in lizards,” *J. Assoc. Res. Otolaryngol.* **12**, 203–217.
- Bergevin, C., Freeman, D. M., Saunders, J. C., and Shera, C. A. (2008). “Otoacoustic emissions in humans, birds, lizards, and frogs: Evidence for multiple generation mechanisms,” *J. Comp. Physiol. A* **194**, 665–683.
- Bergevin, C., Fulcher, A., Richmond, S., Velenovsky, D., and Lee, J. (2012a). “Interrelationships between spontaneous and low-level stimulus-frequency otoacoustic emissions in humans,” *Hear. Res.* **285**, 20–28.



- Bergevin, C., and Shera, C. A. (2010). "Coherent reflection without traveling waves: On the origin of long-latency otoacoustic emissions in lizards," *J. Acoust. Soc. Am.* **127**, 2398–2409.
- Bergevin, C., Velenovsky, D. S., and Bonine, K. E. (2010). "Tectorial membrane morphological variation: Effects upon stimulus frequency otoacoustic emissions," *Biophys. J.* **99**, 1064–1072.
- Bergevin, C., Walsh, E. J., McGee, J., and Shera, C. A. (2012b). "Probing cochlear tuning and tonotopy in the tiger using otoacoustic emissions," *J. Comp. Physiol. A* (in press).
- Bogert, B. P., Healy, M. J. R., and Tukey, J. W. (1963). "The quefrency analysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking," in *Proceedings of the Symposium on Time Series Analysis*, edited by M. Rosenblatt (Wiley, New York), pp. 209–243.
- Choi, Y.-S., Lee, S.-Y., Parham, K., Neely, S. T., and Kim, D. O. (2008). "Stimulus-frequency otoacoustic emission: Measurements in humans and simulations with an active cochlear model," *J. Acoust. Soc. Am.* **123**, 2651–2669.
- Cleveland, W. S. (1993). *Visualizing Data* (Hobart, Summit, NJ).
- Dhar, S., Talmadge, C. L., Long, G. R., and Tubis, A. (2002). "Multiple internal reflections in the cochlea and their effect on DPOAE fine structure," *J. Acoust. Soc. Am.* **112**, 2882–2897.
- Francis, N. A., and Guinan, J. J. (2010). "Acoustic stimulation of human medial olivocochlear efferents reduces stimulus-frequency and click-evoked otoacoustic emission delays: Implications for cochlear filter bandwidths," *Hear. Res.* **267**, 36–45.
- Greenwood, D. D. (1990). "A cochlear frequency-position function for several species—29 years later," *J. Acoust. Soc. Am.* **87**, 2592–2605.
- Harte, J. M., Pigasse, G., and Dau, T. (2009). "Comparison of cochlear delay estimates using otoacoustic emissions and auditory brainstem responses," *J. Acoust. Soc. Am.* **126**, 1291–1301.
- Joris, P. X., Bergevin, C., Kalluri, R., Mc Laughlin, M., Michelet, P., van der Heijden, M., and Shera, C. A. (2011). "Frequency selectivity in Old-World monkeys corroborates sharp cochlear tuning in humans," *Proc. Natl. Acad. Sci. U.S.A.* **108**, 17516–17520.
- Kalluri, R., and Shera, C. A. (2001). "Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation," *J. Acoust. Soc. Am.* **109**, 622–637.
- Kalluri, R., and Shera, C. A. (2012). "Equivalence of swept- and discrete-tone stimulus-frequency otoacoustic emissions," *Assoc. Res. Otolaryngol. Abs.* **35**, 399.
- Kemp, D. T. (1978). "Stimulated acoustic emissions from within the human auditory system," *J. Acoust. Soc. Am.* **64**, 1386–1391.
- Knight, R. D., and Kemp, D. T. (2001). "Wave and place fixed DPOAE maps of the human ear," *J. Acoust. Soc. Am.* **109**, 1513–1525.
- Konrad-Martin, D., and Keefe, D. H. (2003). "Time-frequency analyses of transient-evoked, stimulus-frequency, and distortion-product otoacoustic emissions: Testing cochlear model predictions," *J. Acoust. Soc. Am.* **114**, 2021–2043.
- Lineton, B., and Wildgoose, C. M. (2009). "Comparing two proposed measures of cochlear mechanical filter bandwidth based on stimulus frequency otoacoustic emissions," *J. Acoust. Soc. Am.* **125**, 1558–1566.
- Long, G. R., Talmadge, C. L., and Lee, J. (2008). "Measuring distortion-product otoacoustic emissions using continuously sweeping primaries," *J. Acoust. Soc. Am.* **124**, 1613–1626.
- Meenderink, S. W., and van der Heijden, M. (2010). "Reverse cochlear propagation in the intact cochlea of the gerbil: Evidence for slow traveling waves," *J. Neurophysiol.* **103**, 1448–1455.
- Moleti, A., and Sisto, R. (2008). "Comparison between otoacoustic and auditory brainstem response latencies supports slow backward propagation of otoacoustic emissions," *J. Acoust. Soc. Am.* **123**, 1495–1503.
- Moleti, A., Sisto, R., Paglialonga, A., Sibella, F., Anteunis, L., Parazzini, M., and Tognola, G. (2008). "Transient evoked otoacoustic emission latency and estimates of cochlear tuning in preterm neonates," *J. Acoust. Soc. Am.* **124**, 2984–2994.
- Morlet, J., Arens, G., Forgeau, I., and Giard, D. (1982). "Wave propagation and sampling theory," *Geophysics* **47**, 203–236.
- Neely, S. T., Norton, S. J., Gorga, M. P., and Jesteadt, W. (1988). "Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli," *J. Acoust. Soc. Am.* **83**, 652–656.
- Papoulis, A. (1962). *The Fourier Integral and Its Applications* (McGraw-Hill, New York).
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes in C: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, Cambridge).
- Puria, S. (2003). "Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions," *J. Acoust. Soc. Am.* **113**, 2773–2789.
- Ruggero, M. A., Rich, N. C., Robles, L., and Shivapuja, B. G. (1990). "Middle-ear response in the chinchilla and its relationship to mechanics at the base of the cochlea," *J. Acoust. Soc. Am.* **87**, 1612–1629.
- 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," *J. Acoust. Soc. Am.* **120**, 901–914.
- Shera, C. A. (2003). "Mammalian spontaneous otoacoustic emissions are amplitude-stabilized cochlear standing waves," *J. Acoust. Soc. Am.* **114**, 244–262.
- Shera, C. A., and Guinan, J. J. (1999). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," *J. Acoust. Soc. Am.* **105**, 782–798.
- Shera, C. A., and Guinan, J. J. (2003). "Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning," *J. Acoust. Soc. Am.* **113**, 2762–2772.
- Shera, C. A., 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., Guinan, J. J., and Oxenham, A. J. (2010). "Otoacoustic estimation of cochlear tuning: Validation in the chinchilla," *J. Assoc. Res. Otolaryngol.* **11**, 343–365.
- Shera, C. A., Tubis, A., and Talmadge, C. L. (2005). "Coherent reflection in a two-dimensional cochlea: Short-wave versus long-wave scattering in the generation of reflection-source otoacoustic emissions," *J. Acoust. Soc. Am.* **118**, 287–313.
- Shera, C. A., Tubis, A., and Talmadge, C. L. (2008). "Testing coherent reflection in chinchilla: Auditory-nerve responses predict stimulus-frequency emissions," *J. Acoust. Soc. Am.* **124**, 381–395.
- 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., and Zweig, G. (1993). "Noninvasive measurement of the cochlear traveling-wave ratio," *J. Acoust. Soc. Am.* **93**, 3333–3352.
- 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.
- Sisto, R., and Moleti, A. (2007). "Transient evoked otoacoustic emission latency and cochlear tuning at different stimulus levels," *J. Acoust. Soc. Am.* **122**, 2183–2190.
- Sisto, R., Moleti, A., and Shera, C. A. (2007). "Cochlear reflectivity in transmission-line models and otoacoustic emission characteristic time delays," *J. Acoust. Soc. Am.* **122**, 3554–3561.
- Songer, J. E., and Rosowski, J. J. (2007). "Transmission matrix analysis of the chinchilla middle ear," *J. Acoust. Soc. Am.* **122**, 932–942.
- 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.
- Tognola, G., Grandori, F., and Ravazzani, P. (1997). "Time-frequency distributions of click-evoked otoacoustic emissions," *Hear. Res.* **106**, 112–122.
- Wit, H. P., van Dijk, P., and Avan, P. (1994). "Wavelet analysis of real ear and synthesized click evoked otoacoustic emissions," *Hear. Res.* **73**, 141–147.
- 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.